

2

AIR FORCE

AD-A230 278



HUMAN

RESOURCES

COMPUTER MENU TASK PERFORMANCE
MODEL DEVELOPMENT

Byron J. Pierce

OPERATIONS TRAINING DIVISION
Williams Air Force Base, Arizona 85240-6457

DTIC
ELECTE
DEC 27 1990

Stanley R. Parkinson

Arizona State University
Tempe, Arizona 85282

Norwood Sisson

University of Dayton Research Institute
300 College Park Avenue
Dayton, Ohio 45469

November 1990

Final Technical Report for Period September 1987-February 1990

Approved for public release; distribution is unlimited.

90 12 26 043

LABORATORY

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235-5601

NOTICE

When Government drawings, specifications, or other data are used for any purpose other than in connection with a definitely Government-related procurement, the United States Government incurs no responsibility or any obligation whatsoever. The fact that the Government may have formulated or in any way supplied the said drawings, specifications, or other data, is not to be regarded by implication, or otherwise in any manner construed, as licensing the holder, or any other person or corporation; or as conveying any rights or permission to manufacture, use, or sell any patented invention that may in any way be related thereto.

The Public Affairs Office has reviewed this report, and it is releasable to the National Technical Information Service, where it will be available to the general public, including foreign nationals.

This report has been reviewed and is approved for publication.

DEE H. ANDREWS, Technical Director
Operations Training Division

HAROLD G. JENSEN, Colonel, USAF
Commander

REPORT DOCUMENTATION PAGEForm Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY (Leave blank)		2. REPORT DATE November 1990	3. REPORT TYPE AND DATES COVERED Final Technical Report - Sep 1987 - Dec 1990	
4. TITLE AND SUBTITLE Computer Menu Task Performance Model Development			5. FUNDING NUMBERS C - F33615-90-C-0005 PE - 62205F PR - 1123 TA - 34, 03 WU - 02, 85	
6. AUTHOR(S) Byron J. Pierce Stanley R. Parkinson Norwood Sisson				
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Operations Training Division Air Force Human Resources Laboratory Williams Air Force Base, Arizona 8524C 6457			8. PERFORMING ORGANIZATION REPORT NUMBER AFHRL-TR-90 26	
9. SPONSORING/MONITORING AGENCY NAMES(S) AND ADDRESS(ES)			10. SPONSORING/MONITORING AGENCY REPORT NUMBER	
11. SUPPLEMENTARY NOTES				
12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited			12b. DISTRIBUTION CODE	
13. ABSTRACT (Maximum 200 words) This report provides a review of literature on computer menu interface design and related performance factors. Implications for the design of the user system interface for aircraft simulator instructional support systems are considered. A criterion-based search model that makes predictions as to how the number of alternatives on menu pages affects the search process and the pattern of errors that will result is evaluated. The literature on theoretical and empirical work suggests two additional factors that are recommended for inclusion into the search model: (a) user-perceived relationships among target items sought and menu alternatives available for selection, and (b) the probability of an omission situation where the target item is not subsumed under any of the alternatives available for selection. An experiment was conducted to test the effect that all three of these factors have on menu task performance. Results showed that all three factors significantly influenced menu search and response accuracy. A two-criterion menu model was proposed as a means to explicate the performance results of the menu experiment.				
14. SUBJECT TERMS computer computer-aided instruction interfaces man computer interface teaching methods menu simulation			15. NUMBER OF PAGES 82 16. PRICE CODE	
17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT UL	

PREFACE

This report provides a review of literature on computer menu interface design and related performance factors.

This effort supports the Training Technology objective of the Air Force Human Resources Laboratory (AFHRL) Research and Technology Function by providing a means for enhancing computer/operator interaction and operational ability in computer-aided trainers and instructor/operator station applications.

The work was sponsored by the AFHRL Operations Training Division (AFHRL/OT) and performed under Work Unit 1123-34-02, User/System Interface Requirements, Principal Investigator, Dr. Byron J. Pierce. Additional support was provided by Stanley R. Parkinson, Arizona State University, and Norwood Sisson of the University of Dayton Research Institute under Contract No. F33615-90-C-0005, Work Unit 1123-03-85, Flying Training Research Support, Contract Monitor, Capt Claire Fitzpatrick.

Accession For	
NTIS GRA&I	<input checked="checked" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	



TABLE OF CONTENTS

	Page
I. INTRODUCTION	1
Statement of the Problem	1
User/Computer Communication Techniques	1
II. BACKGROUND LITERATURE	2
Modeling Menu Requirements	2
Empirical Investigations	6
Overview	6
Critique	13
Semantic Information Processing	14
General Evaluation	14
Measurement of Categorical Relationships	15
Synopsis	16
Research Rationale	18
III. METHODS	19
Purpose	19
Determining Semantic Relationships Among Menu Stimuli	19
Menu Set Selection	20
Scaling Requirements	20
Results Summary	22
General Approach	23
Experimental Design	23
Apparatus	24
Subjects	24
Procedure	24
Materials	25
Menu Construction	25
Target-to-Menu Pairings	28
Verifying Assumptions	28
Problem Issues	28
Analytic Procedures	29

Table of Contents (Continued)

	Page
Findings and Implications	29
Omission Trials	30
IV. RESULTS	31
Phase I - Response Accuracy	32
Procedures	32
Correct Alternative Absent Trials	32
Findings	32
Correct Alternative Present Trials	32
Correct Alternative Absent Trials	36
Performance Consistency Across Presentation Modes	38
Phase II - Search Strategies	38
Procedures	38
Findings	41
Synopsis	44
V. THE TWO-CRITERION MENU MODEL	44
Overview	44
Description	45
Data-Fitting Methods	47
Method A	47
Method B	51
Procedures	51
Findings	52
Model Solutions	52
Effects of Menu Design Factors on Model Criteria	53
Synopsis	53
VI DISCUSSION	56
Explication of Menu Task Performance Results	57

Table of Contents (Continued)

	Page
Effects of Variations in Number of Alternatives	57
Effects of Variations in Omission Probability	59
Effects of Variations in Dc	60
Effects of Variations in Di	61
Summary	62
 Research Agenda	 63
 VII. Conclusions	 63
 REFERENCES	 65
 APPENDIX. MODEL PARAMETER, RMSE, AND CORRELATIONS VALUES FOR METHOD A AND B BEST-FIT SOLUTIONS	 67

LIST OF TABLES

Table	Page
1 Possible Menu Decision Outcomes	4
2 Zero Choice Outcomes as a Function of Varying Levels of Number of Alternatives	5
3 Alternative Present Choice Outcomes as a Function of Varying Levels of Number of Alternatives	5
4 Experimental Design	24
5 Menus Constructed from Level III of the Hierarchy	26
6 Menus Constructed from Level V of the Hierarchy	27
7 Comparison of Mean Dc, Di-NN and Di-AVE Values for Correct Alternative Present versus Absent Trials	31
8 Cell-by-Cell Proportions of Hits for Correct Alternative Present Trials	33
9 Cell-by-Cell Proportions of Correct Rejections for Correct Alternative Absent Trials	34
10 Regression Results for Proportion of Hits on Correct Alternative Present Trials	35
11 Regression Results for Proportion of Correct Rejections on Correct Alternative Absent Trials	37
12 Correlations Between Presentation Modes for Arcsine Transformations of Proportions of Hits – Correct Alternative Present Trials Only	39
13 Correlations Between Presentation Modes for Arcsine Transformations of Proportions of Correct Rejections – Correct Alternative Absent Trials Only	40
14 Cell-by-Cell Proportions of Self-Terminating, Exhaustive and Redundant Searches for Correct Alternative Present Trials Where All Subjects Exhibited Determinable Search Strategies	42
15 Regression Results for Search Type Proportions on Correct Alternative Present Trials	43
16 Probability Functions Search Strategies by Response Outcomes for the Two-Criterion Menu Model	48

List of Tables (Concluded)

Table	Page
17 Regression Results for Rapid Rejection Criterion Estimates from Method A and B Solutions	54
18 Regression Results for Rapid Selection Criterion Estimates from Method A and B Solutions	55
A-1 Method A Model Solutions for 7 Correct Alternative Present Equations	68
A-2 Method B Model Solutions for 7 Correct Alternative Present Equations	70
A-3 Method A Summary Statistics for Predicted Versus Observed Proportions Across 32 Cells -- 7 Correct Alternative Present Equation Solutions	72
A-4 Method B Summary Statistics for Predicted Versus Observed Proportions Across 32 Cells -- 7 Correct Alternative Present Equation Solutions	73

LIST OF FIGURES

Figure	Page
1 Menu Hierarchy	7
2 Mean Acquisition Time and Percent Error as a Function of Breadth and Depth	8
3 Response Apparatus Used by Miller	8
4 Search Time as a Function of Depth/Breadth	9
5 Percent Correct as a Function of Depth/Breadth	10
6 Response Time as a Function of Target Type and Organization	11
7 Response Time as a Function of Practice and Target Type for Each Organization	12
8 Hypothetical Distribution of Overall Similarity Values Between Targets and Category Names	17
9 Menu Set Items Derived from Levels III, V, and VI of Miller's Hierarchy	21
10 Areas Defined Under the Distribution Curves for the Two-Criterion Menu Model	45
11 Equations for Computing Areas Under Probability Distributions for the Two-Criterion Menu Model	46
12 Comparison Between Performance Data Results and Model Predictions	58

COMPUTER MENU TASK PERFORMANCE MODEL DEVELOPMENT

I. INTRODUCTION

Statement of the Problem

Ground-based aircrew training devices (ATDs) may be conceptualized as consisting of two main components: the simulation system and the instructional support system (ISS). The simulation component encompasses all that is required to replicate the aircraft and the flight environment. It includes all cockpit, visual and database systems, along with the computer hardware and software required to support them. Most behavioral research conducted on ATDs has focused on issues concerning the required fidelity (stimulus cues) of the simulation.

The other component, the ISS, provides all interface controls and displays required to use the simulator as a training device. The ISS also can provide briefing and debriefing capabilities, as well as a variety of other instructional support functions. With advances in visual simulation, computer, and networking technologies, the training capabilities of ATDs have increased dramatically. These advances have increased the complexity of ISS functions and, as a result, the communication between the user and the ISS has become more difficult.

The negative effect that poorly designed ISS interfaces have had on the use of ATDs has been well documented. The United States Air Force has completed a series of research efforts to assess the effectiveness of operational ATDs in the Air Force inventory (Polzella, Hubbard, Brown, & McLean, 1987; Semple, Cotton, & Sullivan, 1981). These studies examined the utilization of these training devices, the use and value of ISS features, and the overall acceptance of the training systems by the students and instructors who used them. The results revealed that many ATDs were not being used as expected, that user acceptance was sometimes low, and that many trainer capabilities were not being utilized. These studies suggested that much of the problem can be attributed to the fact that information displays and controls required to employ ATD capabilities effectively are often either not available or configured in a manner that is difficult to use. These findings provided the genesis for the work described in this report.

User/Computer Communication Techniques

Many practical alternatives exist for communication between the computer and its user. All have certain advantages and disadvantages, which usually are dependent on the characteristics of the typical user of the system (Hauptman & Green, 1983). Menu design has been found especially effective for systems that have new, infrequent or untrained users, because the user need not memorize and recall commands but merely recognize them. Thus, effective menu design has become an area of concern for designers of simulator ISSs where the end user is typically an instructor or a student pilot who uses the system on a recurring but infrequent basis. Via menus, whenever the user must control computer actions the program displays a set of options describing all the alternatives available at that point. The user selects one of the alternatives by keying-in a response that activates the option desired. The program then branches to the subroutine that corresponds to the user's choice and displays a new menu of available options. This process is repeated until the user finds and selects the end-level option sought.

The primary advantage of a menu-driven system is that the user has only to know and understand the current options available at any particular point of execution. Little formal training is required because this technique relies mainly on the operator's general subject-matter expertise and requires only passive responses to computer prompts. The main drawbacks associated with menu-driven systems stem from the need to organize all available command options into

a particular menu structure within the program. Thus, the creator of the menu not only must know all the possible or desirable options to include, but also must organize this material into a menu configuration that accommodates a wide variety of users.

Even though the utility of menu-driven systems to ISS applications has been well accepted, a need exists for expanded understanding of the performance effects of menu design factors. Such information could be used to structure menus so as to optimize user performance. Menu design factors which have been found to impact user performance include breadth (number of items displayed on each menu level), depth (number of levels within a menu structure), and item organization. The objective of the present investigation was to quantify and model the effects these factors have on user performance at rudimentary levels of task familiarity. The ultimate goal was to build a theoretical foundation which can be applied specifically to the design of menus for ISSs.

II. BACKGROUND LITERATURE

In very simple applications where the number of functions performed by a system are relatively few, menu commands can usually be arranged on a single page from which the user selects the command desired. In more complicated applications where the system is capable of performing numerous functions, a more sophisticated organization of commands is required. Large menus typically are organized into hierarchies with varying levels of menu pages and options. These hierarchies can be arranged with many items on a menu and a minimum number of sequential menu levels (breadth), with few items on each menu and several levels (depth), or with some intermediate level of breadth and depth. Recommendations concerning the use of breadth versus depth in menu construction have been based on quantitative and empirical studies, as well as implications derived from cognitive theories. What follows is a general review of experiments and theories that are pertinent to the design of computer menu interfaces.

Modeling Menu Requirements

Only a limited number of quantitative approaches for resolving menu breadth/depth issues have been investigated. Lee and MacGregor (1985) described a mathematical model for calculating optimal menu breadth as a function of human and computer factors of search strategy, scanning time, keypress time and computer response time. The model was based on the assumption that the usefulness of a menu depends on the amount of time required to retrieve information associated with its use. Thus, the model calculates optimal breadth by minimizing the total time required of the user to navigate through a menu and select the target (i.e., end-level item) desired. Model solutions are dependent upon the search strategy to be employed by the user. Possible search strategies include: self-terminating, wherein the user terminates the search process as soon as the selected alternative is encountered in the menu; exhaustive, wherein the user reads all alternatives in the menu before responding; and redundant, wherein the user examines several or all alternatives more than once before responding.

Lee and MacGregor (1985) used this model to calculate the optimum number of alternatives for reading rates, keypress times and computer response times typical of menu-based videotex information retrieval systems. Videotex is an interactive system designed to present textual and graphic information at the command of the user (Tydeman, Lipinski, Adler, Nyhan, & Swimpfer, 1982). The user typically employs a keypad or keyboard and a computer terminal to interact with a central computer via telephone lines or hardware connection. Videotex systems involve menus which may consist of thousands of menu pages with menu alternatives that are frequently phrases or lists. For the range of values examined, the authors concluded that 4 to 8 alternatives per page were optimal to minimize user search time.

MacGregor, Lee, and Lam (1986) refined this model and examined those factors which directly impact upon the search processes employed by users. They pointed out that for conditions where computer response time is long and the decision time short, the optimal breadth can exceed the "4 to 8" range that Lee and MacGregor (1985) recommended. Such conditions are typical of command-type menu applications. By comparison to videotex systems, command-type menus involve a relatively limited set of alternatives, where learning of items and their location in a menu can occur quite rapidly. Command items themselves are frequently single words requiring little search and processing time. MacGregor et al. (1986) argued that the greater menu breadth required by command menu applications is predicted by the search model and that the "4 to 8" range previously suggested was based on parameter values typical of videotex systems.

Although the factors which contribute to user response times may range in value from one type of menu-based application to another, MacGregor et al. (1986) argued that the search model is applicable across conditions. To date, however, neither this model nor the empirical literature provides a clear understanding of how to derive an optimal structure given a particular menu application. To increase the usefulness of the Lee and MacGregor (1985) search model, the majority of the factors which account for search/decision and response times must be identified, and procedures for their measurement must be developed and applied to specific applications of interest.

MacGregor et al. (1986) made progress toward this end by examining factors which influence user decision processes for videotex menu retrieval. The authors hypothesized that the decision process used during menu retrieval is criterion-based. The criterion is equivalent to that level of probability/confidence which an alternative must exceed to be considered as a choice. Menu search processes were seen as a manifestation of the same decision process. A self-terminating search occurs when an alternative clearly exceeds the criterion and is chosen without further search. If an alternative exceeds the criterion by a smaller amount, or if several alternatives exceed the criterion level, an exhaustive or redundant search must be undertaken.

The authors stated that the number of alternatives presented on the menu page directly influenced the criterion level used during the decision process. They argued that as number of alternatives (a) increases, the lower bound of the criterion ($1/a$) decreases. From this criterion-based search model, several predictions were generated concerning the effects that variations in number of alternatives have on user search strategies and on the frequencies of various types of errors.

MacGregor et al. (1986) examined the validity of model predictions in an experiment that incorporated a partial search procedure in which subjects examined only one of p menu pages from a videotex-type menu hierarchy and selected an alternative on that page. This method provided the capability of examining the decision and search process, as well as permitting the empirical estimation of parameter values for the model. Alternatives were presented either simultaneously or sequentially. In the simultaneous condition, all the alternatives were viewed at one time, in the normal fashion. In the sequential condition, the initial display showed only the set of numbers corresponding to menu choices so that subjects knew how many alternatives there were. Thereafter, they could display only one alternative at a time, with the sequence and timing of each exposure being controlled by the subject. The sequential presentation procedure permitted the study of decision and search processes employed by subjects. The simultaneous condition was used for comparison purposes to determine whether subject performances in the sequential condition were influenced by the forced sequential search. A mixed design was used, with one between-groups factor and one within-group factor. The between factor was sequential versus simultaneous presentation of alternatives. The within factor was the number of alternatives per page: either 2, 4, 8, or 16.

To a considerable extent, the results supported the predictions. First, it was predicted that no subject would use any single strategy but rather, would employ different strategies. The

results indicated (a) that no subject followed one strategy to the exclusion of the others, and (b) that two-thirds of the subjects used all three strategies. The model also predicted that both self-terminating and redundant searches would increase as the number of choices per page increased. The results indicated some support for these predictions. When the number of alternatives were few, exhaustive searches predominated. As the number of alternatives increased, both redundant and self-terminating searches tended to increase in frequency. However, although self-terminating searches continued to increase up to 16 alternatives per page, exhaustive searches also increased and redundant searches dropped off, contrary to predictions. The authors suggested that this may have been due to the effort involved in reading pages with so many alternatives. An alternate explanation, which will be examined in greater detail later in this report, is that the results may have been confounded by the variations in semantic relatedness among the targets and alternatives across conditions.

Another problem with these results concerns the manner in which exhaustive and self-terminating searches were operationally defined and analyzed. The authors defined exhaustive searches as occurring when each alternative was examined once and once only. Self-terminating searches were defined as occurring when less than all alternatives were examined prior to subject response. Based on these definitions, when subjects examined all alternatives once and only once and then chose the alternative in the last menu position, this search was considered exhaustive. It could be argued that such cases inflated the frequencies of exhaustive searches reported in the MacGregor et al. (1986) study and conversely deflated the reported frequencies of self-terminating searches. As the number of alternatives decreased, the percentage of these cases increased due to the increased probability of having the correct alternative occupy the last position in the menu. Therefore, it is reasoned that the degree of the inflation of exhaustive searches and deflation of self-terminating searches in the results reported by MacGregor et al. (1986) was inversely related to number of alternatives. Thus, the reported increase in self-terminating searches and decrease in exhaustive searches as a function of increasing number of alternatives may have been confounded by the researchers' operational definitions of search strategy.

The model made a number of predictions concerning subject decisions and the incidence of different types of errors. There are only two possible correct answers in a menu search: Either the correct alternative is present in the menu and is chosen (which defines a "hit" outcome); or the correct alternative is not present in the menu and the user makes a zero choice, indicating that none of the alternatives is correct (which defines a "correct rejection"). There are three possible incorrect subject responses: The subject can choose an incorrect alternative when the correct alternative is present (which defines a "commission miss"); the subject can choose an incorrect alternative when the correct alternative is not present (which defines a "false alarm"); or the subject can make a zero choice when the correct alternative is present (which defines an "omission miss"). Table 1 summarizes the possible menu decision outcomes under this paradigm.

Table 1. Possible Menu Decision Outcomes

User choice	Correct alternative present	Correct alternative not present
User chooses the Correct Alternative	Hit	Not Possible
User chooses an Incorrect Alternative	Commission Miss	False Alarm
Zero Response (user indicates correct alternative is not present in menu)	Omission Miss	Correct Rejection

The model predicted that the number of alternatives would inversely affect the subjects' frequency of rejecting all alternatives and, conversely, that the incidence of selecting some alternative would be related directly to number of alternatives. Given that a zero choice was made, the model further predicted that the conditional probability of an omission miss would decrease as a function of number of alternatives and the probability of a correct rejection would increase. The results tended to support these predictions (Table 2). Significant effects due to number of alternatives accounted for about 15% of the variance in each analysis. The model also predicted that hits would be related inversely to number of alternatives and that commission misses would be related directly. Both hits and commission misses showed some trend in the predicted directions, but the 2- and 4-alternative conditions showed a reversal in trend (Table 3). In both cases, the significant effect due to number of alternatives accounted for about 46% of the variance.

Table 2. Zero Choice Outcomes as a Function of Varying Levels of Number of Alternatives (from MacGregor et al., 1986)

Outcome	Number of alternatives/page			
	2	4	8	16
Frequency of zero choice	53	47	27	36
Probability of omission errors	0.45	0.40	0.15	0.14
Probability of correct rejections	0.55	0.60	0.85	0.86

Table 3. Alternative Present Choice Outcomes as a Function of Varying Levels of Number of Alternatives (from MacGregor et al., 1986)

Outcome	Number of alternatives/page			
	2	4	8	16
Frequency of choosing an alternative	187	193	213	204
Probability of commission errors	0.27	0.09	0.34	0.39
Probability of false alarms	0.10	0.10	0.12	0.08
Probability of hits	0.63	0.81	0.55	0.53

To a large extent, the findings of MacGregor et al. (1986) supported their contention that the search process can be viewed usefully as a criterion-based decision process. The results provided evidence that the number of alternatives presented on a menu page affects the level to which this criterion adjusts, and consequently influences the search strategy, search times, and the types of errors that occur. The authors suggested that because some components of error were a direct function of number of alternatives and others were an indirect function, an implication of the model was that, overall, errors would sum to a quadratic function of the number of alternatives, with intermediate levels leading to optimal performance. The results supported this prediction, with 4 to 5 alternatives providing the best arrangement.

A major test of the validity of this search/decision model rests on how well it predicts user performances across a variety of conditions. MacGregor et al. (1986) stressed that parameter values which resulted from their study were pertinent to the use of menus in large information retrieval applications and would not likely apply to the daily user of a command-type menu. They suggested that the experienced user of a command menu is likely to learn not only the correct alternatives for a given target item but also the positions of these alternatives on the page, so that only a small subset of alternatives need be scanned at high rates of reading time. The values for the model parameters under these conditions would be quite different from those computed from the MacGregor et al. (1986) study, yielding an optimal structure with less depth and more breadth than the one recommended for videotex menus.

At issue is the MacGregor et al. (1986) contention that even though the parameter values may change, the basic search model and its predictions should hold across a variety of menu applications. The question arises as to whether the processes involved during command menu search are the same as those suggested for videotex menus. More specifically, if the optimum number of alternatives can be determined by the same criterion-based processes in both videotex and command menus, then are variables which affect the decision criterion the same in both conditions? To the extent that this is true, then both the theoretical and empirical literature suggest factors in addition to the number of menu alternatives that may play important roles in the menu search and decision process for users at initial levels of menu structure familiarity. These factors include the cognitive relationships among targets and alternatives, and the probability that the target is not subsumed under any available alternative (i.e., omission trials). What follows is a review of issues from the empirical literature relevant to these factors. This review forms the basis for recommending that the search model be modified in such a manner that it accounts for performance effects of these additional factors.

Empirical Investigations

Overview

The need for an empirical resolution of the breadth/depth issue led Miller (1981) to investigate the influence breadth and depth had on the speed and accuracy of menu selection performances. Miller (1981) had college students search for a word presented in one of four hierarchical structures: (a) menus with 2 alternatives at each of six levels (2^6), (b) menus with 4 alternatives at each of three levels (4^3), (c) menus with 8 alternatives at each of two levels (8^2), and (d) menus with 64 targets at a single level (64^1). Each hierarchy had the same 64 targets at its lowest level (see Figure 1). Targets were drawn from eight basic categories, with eight targets per category. Each of the hierarchies was based on subordinate and superordinate organization of the items contained in each of these basic categories. For the broadest menu (64^1), subjects were required to search the array for the target word and respond by pressing a button corresponding to the location of a randomly arranged group of words containing the target. For deeper menus, subjects were shown names of categories that might contain the target and were asked to select a sequence of options which would lead to the target word. Once the target was found, subjects responded by pressing a button which identified the location of the target on the display. Miller (1981) found that both speed and accuracy varied with menu structure. Subjects were slower with the extreme levels of breadth and depth (i.e., the 64^1 and 2^6 menus - see Figure 2). Error data also produced a U-shaped function, with the least amount of error data produced by the 8^2 menu configuration. Overall, Miller's (1981) findings would suggest the use of structures of intermediate breadth and depth in menu design.

Snowberry, Parkinson, and Sisson (1983) pointed out two potential shortcomings of Miller's (1981) methodology. The first concerned the arrangement of items used in the broadest menu structure (i.e., the 64^1 menu). Though all other menus were constructed with categories intact, the 64^1 menu had targets arranged in a random fashion. Thus, breadth in the broadest menu seemingly was confounded by the omission of strict categorical grouping of display options. The second criticism levied against Miller's (1981) methodology concerned response requirements. In all but the broadest menu, stimuli were presented along the right and left edges of the display, laterally adjacent to response buttons mounted on the side edges of the display (see Figure 3). Subjects input their responses by pressing the button adjacent to the alternative chosen. In the 64^1 menu condition, eight pushbuttons were mounted along the top and bottom of the display, adjacent to the eight columns of items which contained all the target words. Due to physical limitations of the apparatus, subjects in this condition selected the block of eight words containing the target word, rather than identifying the actual word itself. Snowberry et al. (1983) contended that because breadth and response type were confounded, one should not draw the conclusion that the slower search time in the broadest menu condition resulted from breadth.

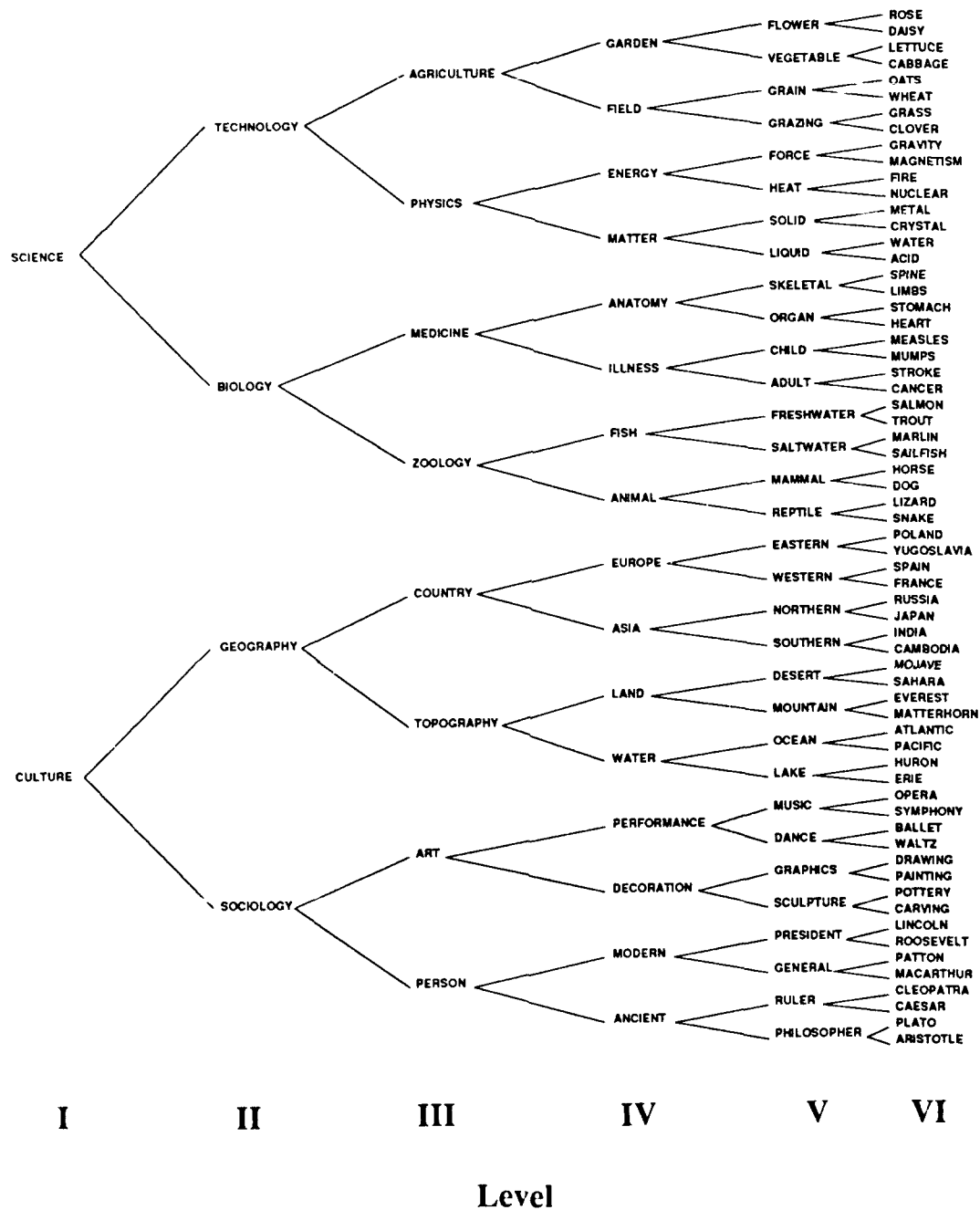


Figure 1. Menu Hierarchy (adapted from Miller, 1981, and Snowberry, Parkinson, and Sisson, 1983).

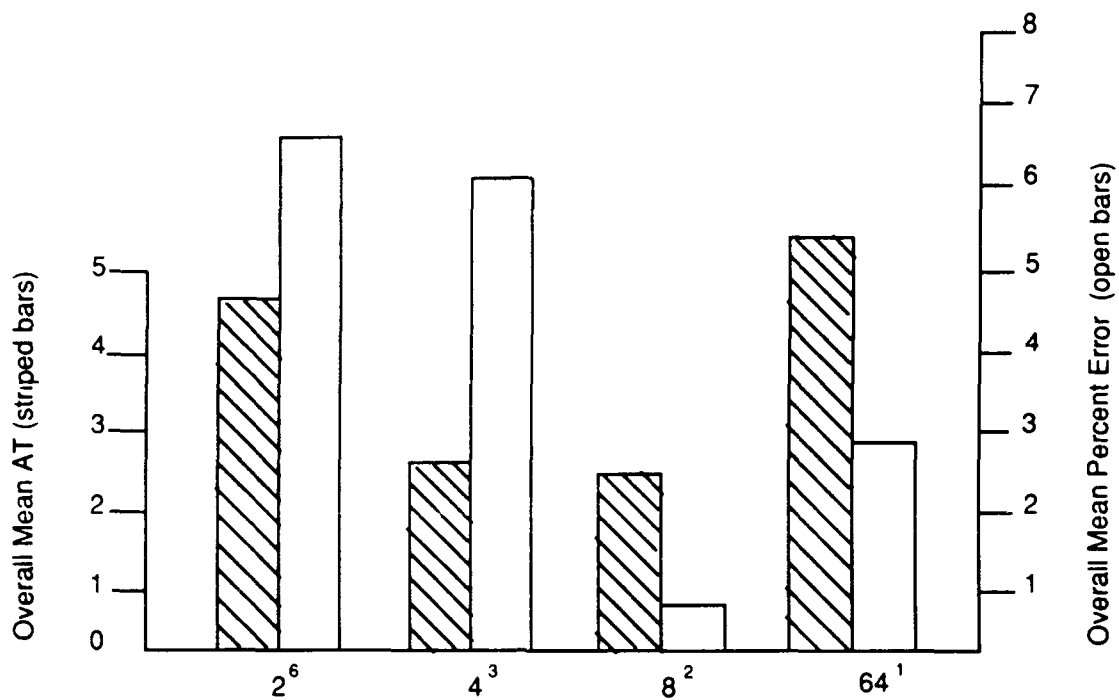


Figure 2. Mean Acquisition Time (AT) and Percent Error as a Function of Breadth and Depth (adapted from Miller, 1981).

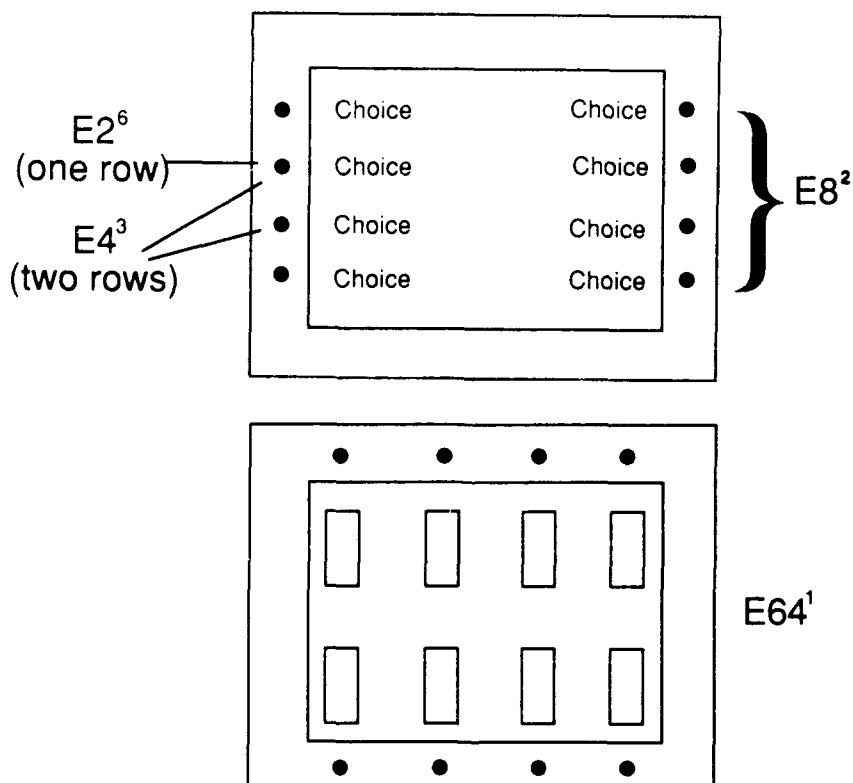


Figure 3. Response Apparatus Used by Miller (1981) (adapted from Miller, 1981).

In a study designed to resolve these issues, Snowberry et al. (1983) used both a randomized and a categorized 64-item display to compare performances on menus of increasing depth and decreasing breadth. Response requirements were held constant across conditions by requiring subjects to respond by keying-in a two-digit reference number appearing on the display directly adjacent to the item options available. Both targets and hierarchical structures were identical to those used by Miller (1981), as shown in Figure 1. Comparison of the random organization 64¹ menu condition to other menu breadth/depth conditions produced response time results similar to those obtained by Miller (1981). Specifically, comparison of response times across varying conditions of breadth and depth revealed that subjects were faster with intermediate breadth and depth levels and slower with extreme menu levels. When category organization was held constant across conditions, however, search time improved as depth decreased and breadth increased (Figure 4). In addition, contrary to Miller's (1981) findings, accuracy was found to improve as breadth increased (Figure 5). This result held regardless of the display organization (random versus categorized) used in the 64¹ condition.

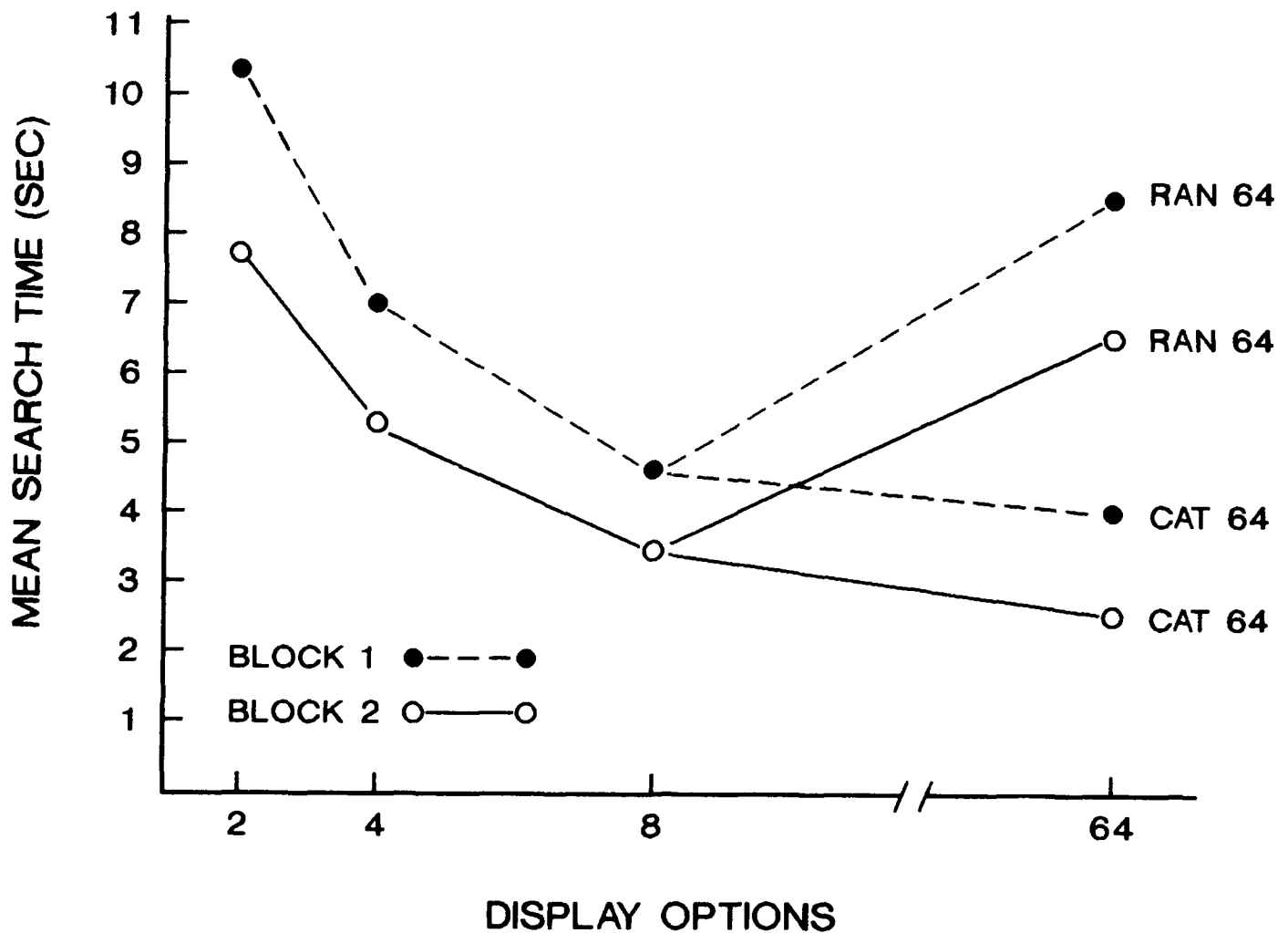


Figure 4. Search Time as a Function of Depth/Breadth
(adapted from Snowberry et al., 1983).

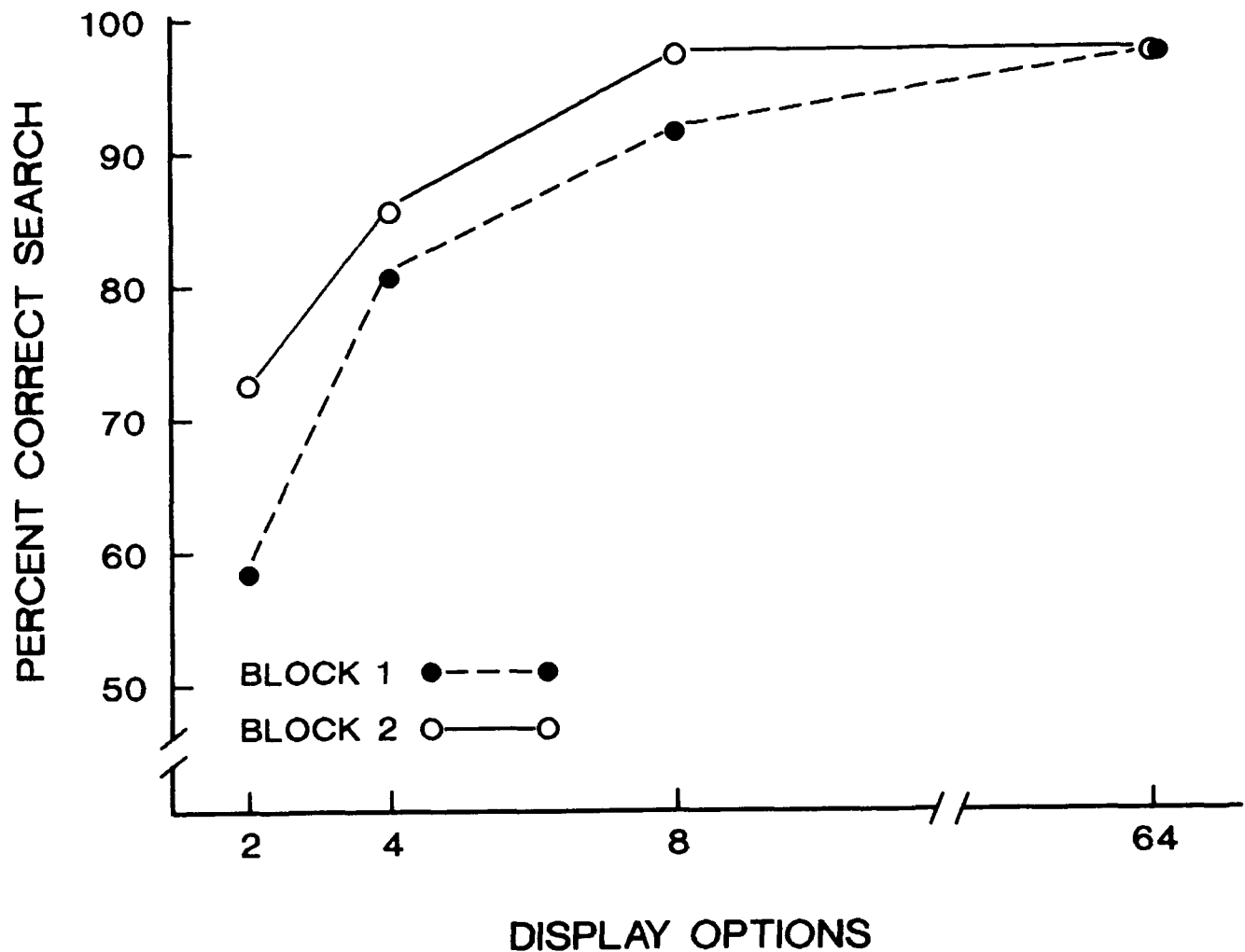


Figure 5. Percent Correct as a Function of Depth/Breadth
(adapted from Snowberry et al., 1983).

The organization of items within a menu was shown to have significant consequences on performances by the subjects for the 64¹ menu condition in the Snowberry et al. (1983) study. Other studies also have addressed this issue. For example, Card (1982) found that for menus of 18 items there was an advantage for an alphabetical menu organization over categorical and random orderings. However, McDonald, Stone, and Liebelt (1983) contended that this finding may be applicable only when users do not have a well-formed cognitive organization of task categories.

In an attempt to support their contention, McDonald et al. (1983) examined the effects of alphabetical, categorical and random organizations on search performances using broad (64-item) menus. They examined five types of menu organizations in which four lists (i.e., columns) of 16 items each were combined to form the 64-item menus. In three of the menus, each column contained a different category list of items. Within these lists, items were arranged categorically (CC), alphabetically (CA), or randomly (CR). Within-list categorical orderings (menu CC) were based on subjects' similarity ratings for pairings of items contained in each category list. These ratings were then subjected to a multidimensional scaling and hierarchical clustering technique from which categorical orderings were produced. Of the remaining two menus, one had all 64 items arranged in a completely alphabetical order (A) while the other had all items randomly assigned (R) across lists. In addition to the categorization conditions, the effect of target type was assessed by comparing performance using explicit targets as stimuli to performance obtained

using target definitions. Thus, a 5x2 factorial design was used which contained five levels of the category condition and two levels of target stimuli.

Subjects were hired from a temporary employment agency as general secretarial help. The subjects' task in this study was to locate an item in the menu and enter its identification letter on the computer keyboard. Results showed that response times were generally faster with explicit targets than with definitions, particularly during the initial session of the five-session study (see Figures 6 and 7). With both target types, categorically arranged menus (menus CC, CA and CR) produced the shortest response times, while the random organization of items across lists (menu R) produced the longest times. Separate analyses were conducted on the response times for the first session of trials, where it was reasoned that learning effects were at a minimum. Significant main effects were found for both item organization and target type. Within-list categorical organization (menu CC) produced the shortest response times, as did the use of explicit targets. Finally, when target definitions were used, all categorically arranged menus (menus CC, CA and CR) showed shorter response times than those for the alphabetical organization (menu A). However, with explicit targets, alphabetical organization (menu A) produced response times almost as short as those produced by within list categorical organization (menu CC).

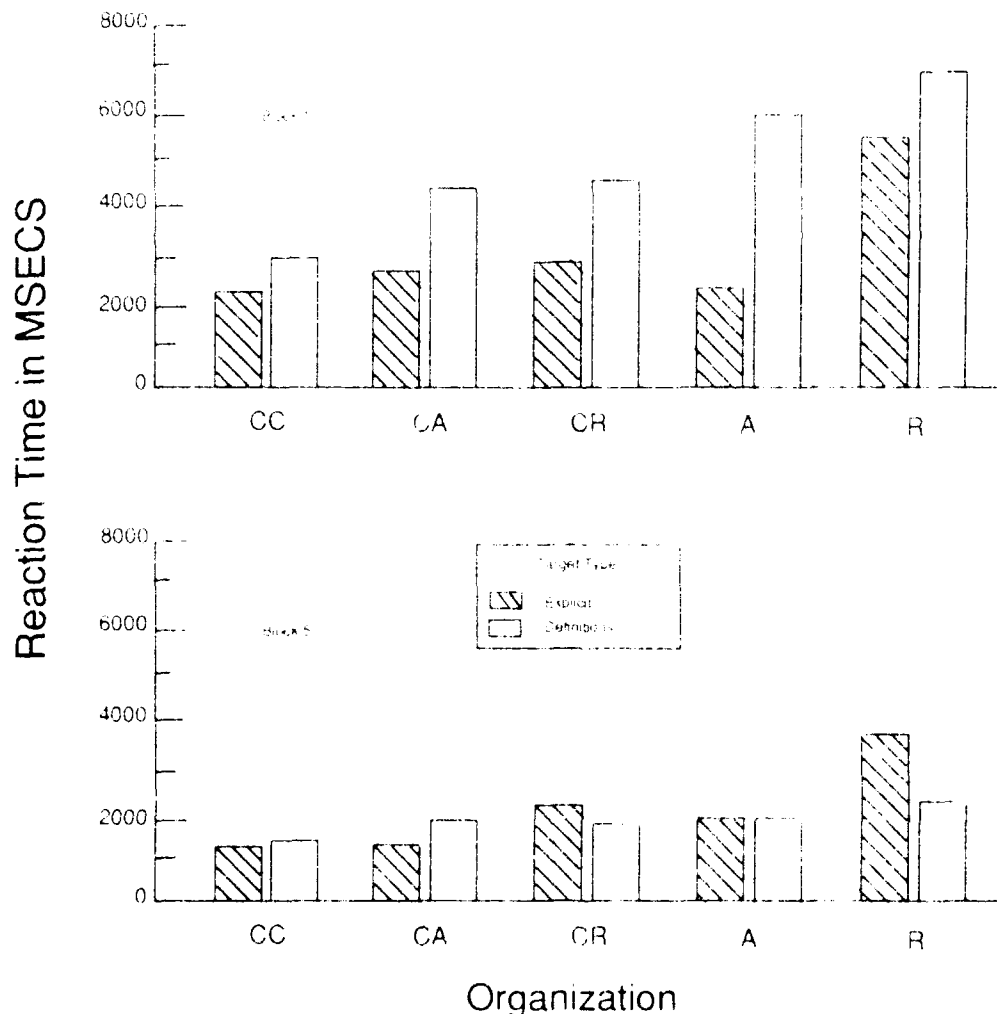


Figure 6. Response Time as a Function of Target Type and Organization (adapted from McDonald et al., 1983)

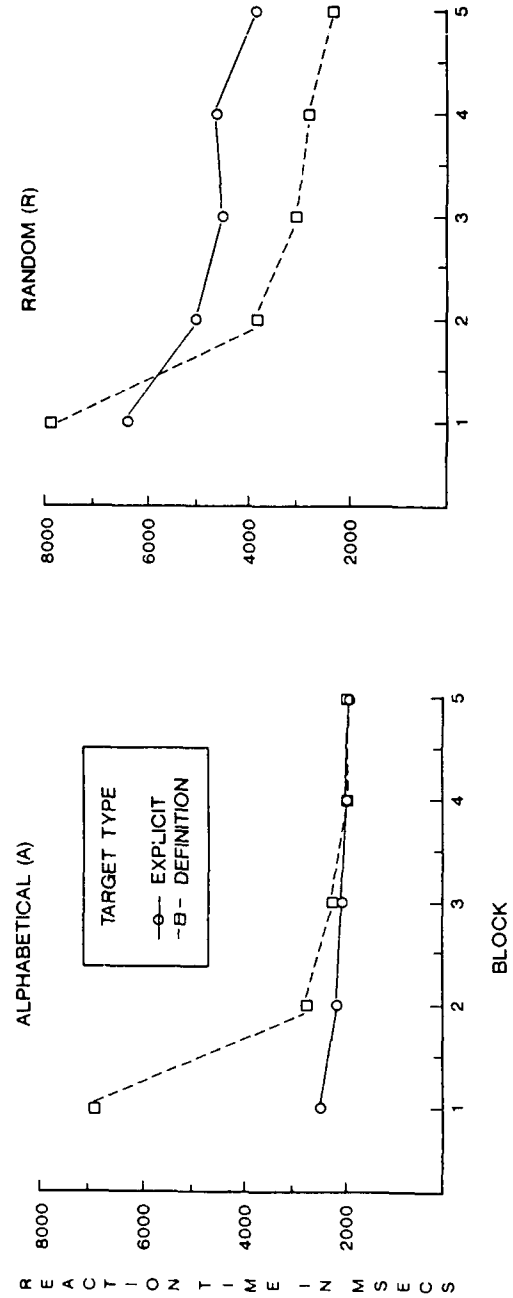
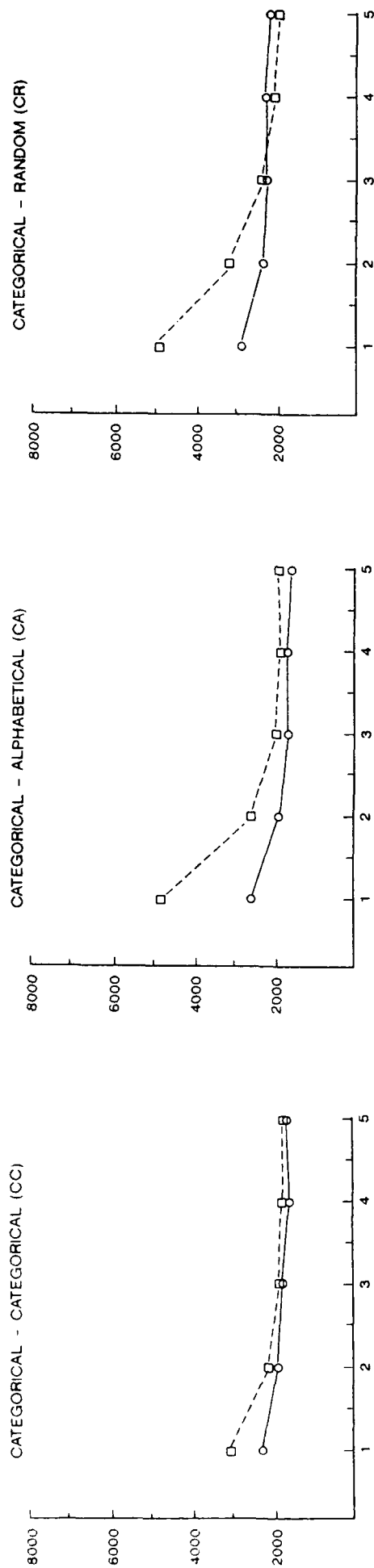


Figure 7. Response Time as a Function of Practice and Target Type for Each Organization
(adapted from McDonald et al., 1983).

McDonald et al. (1983) interpreted the results as showing the superiority of a categorical menu organization over a purely alphabetical one. This effect was evident particularly in situations where there was uncertainty about the target. However, the results did not support the authors' contention that it is possible that other organizations, particularly alphabetical ones, might be superior when users do not have a well-formed cognitive organization for the task domain. Under no circumstances did alphabetical organization (menu A) produce performances superior to performance resulting from the within-list categorical organization (menu CC). Even in the first block of trials when mental organizations of the task domain should be poorest, performance in the CC condition was superior to condition A performance. Thus, the validity of the McDonald et al. (1983) contention concerning conditions under which various organizational types might produce superior performances remains in question.

Critique

The studies conducted by Card (1982), Snowberry et al. (1983), and McDonald et al. (1983) all support the contention that for command-type menu systems, optimal menu breadth is somewhat larger than that which MacGregor et al. (1986) recommended for videotex menus. In addition, all these studies showed significant effects of menu categorization on retrieval time for broad menu structures. MacGregor et al. (1986) contended that grouping helps for precisely the same reason that hierarchical organization helps: It restricts the options that have to be searched. In either case, however, search is restricted only to the extent that the structure reflects the user's perceived relationships among targets and menu alternatives. For frequent users of a system, learning the menu structure can be aided by designing the system to reflect preconceived menu item relationships. For infrequent system users, designing the system to reflect user perceptions of these relationships becomes even more critical.

The effect that user-perceived relationships among targets and menu alternatives has on performance has received little empirical attention. In simple command menus of the type used by Miller (1981), Card (1982), Snowberry et al. (1983), and McDonald et al. (1983), cognitive relationships might best be thought of in terms of semantic similarity or distances among menu items. The lack of attention to and, perhaps more importantly, failure to control this factor may represent a possible confounding of the findings in several menu optimization studies. In support of this argument are semantic memory models (e.g., Collins & Quillian, 1969; Rips, Shoben, & Smith, 1973) which contend that the semantic distance between a subset category and its member is closer than the distance between the superset category and the member (e.g., robin and bird versus robin and animal). Thus, the results of breadth/depth studies in which semantic distances between targets and category sub/supersets were not controlled may have been confounded by this factor. For example, recall that both Miller (1981) and Snowberry et al. (1983) varied breadth and depth by manipulating subordinate and superordinate organization of 64 targets drawn from eight basic categories (review Figure 1). As menu depth varied, so did the semantic distances from the target to the menu alternatives. To the extent that semantic distance variations may have affected performances, performance variations that were attributed to breadth/depth manipulations may have been confounded.

Another factor not formally considered in the search model is the effect of variations in omission probability; that is, the probability that the target is not subsumed under any of the available alternatives. For discriminate decision processes, it is known from signal detection theory that as probability of an omission trial varies, decision criterion levels shift (Green & Swets, 1966). If, as suggested by MacGregor et al. (1986), the search model involves a criterion-based decision process, then omission probability should have a significant effect on user criterion levels.

The above discussions regarding semantic relationships among menu items and omission probability lead to the following hypotheses concerning processes involved in the search model:

(a) For simple command-type menu structures, semantic relationships among targets and alternatives can significantly influence the search and decision processes involved in menu task performance; and (b) similar to the manner in which variations in the number of menu alternatives affects the decision criterion, omission probability also significantly affects the decision criterion used in the decision process. The concept of omission probability is relatively straightforward, but conceptualizing and measuring semantic relationships among targets and alternatives are more complex. Theoretical models and empirical findings, in the areas of categorization and the organization and structure of semantic memory, provide clues as to how to best address this problem. The next section provides a general review of this material, along with implications concerning the definition and measurement of menu item relationships.

Semantic Information Processing

The separation of long-term memory into episodic and semantic memory components was first suggested by Tulving (1972). He conceptualized two separate but interrelated systems for the storage of differing types of information: The episodic component stored temporal, sequential information concerning instances and events, as well as biographical information; the semantic component was defined as memory for language and language syntax. A general evaluation of research and theoretical positions concerning the nature of the semantic memory component is given below.

General Evaluation

The primary findings from studies which have focused on the processes involved in categorization and the organization of semantic memory can be summarized as follows:

1. Effects of graded structure of category members. A central issue in categorization research has been the finding that categories possess graded structure. Graded structure refers to the degree to which members provide a good example of their various category classes. For a given level of concepts within a semantic hierarchy, graded structure has been shown to be strongly related to both the speed and accuracy of verifying class membership (Rips et al. 1973; Rosch, 1973, 1975; Shoben, 1976; Smith, 1978; Smith, Balzano, & Walker, 1978).

2. Effects of graded structure of category nonmembers. Semantic relationships between categories and nonmembers are graded in much the same manner as are members of a category. Graded structure has been shown to affect both speed and accuracy in verifying nonmembers of a class (Barsalou, 1983; Shoben, 1976; Smith et al., 1978).

3. Category size effects. Both speed and accuracy in semantic verification tasks have been shown to be significantly related to the level at which the category resides within a network hierarchy (Collins & Quillian, 1969, 1970). The reversal of this effect can occur for category structures in which subordinate-level concepts have lower semantic similarity to exemplars than do superordinate levels (Rips et al., 1973; Smith et al., 1978).

A typical menu navigation task can be conceptualized in terms of the multiple verification of category relatedness of individual alternatives presented on a menu page to a particular target. Although no studies could be found in the literature which bear directly on the problem of defining those semantic variables which affect performance on such a task, the above findings from semantic verification studies can be generalized to the multiple verification problem in an attempt to provide a framework for considering decision processes that may be involved in menu task performance. The varying semantic relationships between the target and individual alternatives on a menu page can be expressed in terms of the graded structure of category

members and nonmembers, as well as each alternative's level within the menu hierarchy. An implication from the above findings is that menu item selection will at least partially depend on the semantic similarity between the target and the set of alternatives presented on the menu page.

To the extent to which the relationships among targets and menu set alternatives are graded both within and across levels of the menu hierarchy, the problem of evaluating their effects on menu task performance is reduced to one of measuring these varying relationships by empirical study. The next section provides an overview of techniques that have been employed as measures of graded structure and semantic similarity in categorization and semantic memory research.

Measurement of Categorical Relationships

The requirements of the present investigation called for measures which reflect the graded relationships among targets and menu alternatives. Research in this area has shown that, to a large extent, the graded structure of an exemplar can be determined in terms of its degree of similarity to other category members and its degree of dissimilarity to category nonmembers. These findings suggest that measurement requirements for the present work can be satisfied by deriving valid measures of semantic similarity for targets with respect to a particular set of alternatives presented on a menu page for selection.

Measures of semantic similarity are usually obtained by having subjects make pairwise comparisons of all items in a stimulus set or of a subset of these item combinations and provide similarity ratings based on subjective evaluation. Ratings are then averaged across subjects and used as predictors of response times and error rates on experimental tasks. Rips et al. (1973) introduced this procedure for use in predicting semantic verification response times and it has subsequently been used in several studies of semantic memory (e.g., Barsalou, 1985; Rosch, 1975; Shoben, 1976).

In addition to demonstrating the significant predictive capability of raw rating averages, both Rips et al. (1973) and Shoben (1976) demonstrated the predictive potential of semantic distances derived from the multidimensional scaling (MDS) of subjects' similarity ratings. MDS takes as input the semantic similarity ratings of all possible pairings of the stimulus set, and yields as output an Euclidean solution in n dimensions where the orientation of these dimensions is fixed by the scaling procedure itself. The advantages of the use of the MDS-derived distance values over raw rating averages are both substantive as well as methodological.

The substantive contribution of MDS solutions is that they offer a method for recovering underlying structural features of scaled items. This may be achieved through the identification and interpretation of directions or axes through the spatial representation. Thus, moving farther along some particular direction, points are successively encountered which correspond to objects that possess more and more of some particular, identifiable property. Such results have been used to describe the characteristic features of category terms in semantic memory studies (e.g., Rips et al., 1973; Shoben, 1976).

The methodological contribution of MDS is twofold. First, MDS provides a less noisy measure of relatedness than do raw rating averages. Noise here refers to error that distorts the measurement of the relationships among items in the stimulus set. MDS proponents claim that MDS provides a more meaningful and interpretive configuration by smoothing out much of this noise through its computations. For example, both Rips et al. (1973) and Shoben (1976) demonstrated that in certain semantic memory tasks, the Euclidean distances derived from MDS surpassed the raw rating averages in predicting subject performances. The authors argued that

the MDS solutions reduced the noise in the data by forcing the configuration into only the few critical dimensions which were reflective of feature similarity among stimuli.

The other methodological advantage that certain MDS techniques hold over raw rating averages concerns the metric properties associated with MDS. MDS can be thought of as a means of recovering metric configurations from nonmetric rating information. The techniques used in MDS assume that within a matrix of rating sums there exists a true underlying configuration of points in n-dimensional Euclidean space that can be ascertained by the linear ordering of interpoint distances (Kruskal, 1964). MDS procedures are designed to recover this underlying configuration. Though the nonmetric properties of raw ratings tend to make questionable the use of certain types of arithmetic operations, the metric properties associated with the MDS configuration enable the use of parametric operations without fear of assumption violation.

For example, later in this report, a technique will be detailed which attempts to combine two separate sets of pairwise similarity distance values based on distance values between items common to each set. The technique produces a single configuration containing all items from both sets through the scaling, translation and reorientation of items in the second set into maximal congruence with common items in the first set. Although the operations involved in this transformation might be questionable for use with nonmetric data, they are perfectly justifiable for use with MDS solutions.

The substantive advantages of MDS are not particularly relevant to the present work. The main interest here is in obtaining valid measures of the semantic relationships among sets of items. Thus, the identification of underlying dimensional characteristics is not of real consequence. The methodological advantages of MDS, however, are of significant interest. Particularly, the metric properties associated with MDS solutions provide a substantial advantage over raw rating averages. Therefore, the decision was made to employ distance values derived from MDS analysis of rating averages as measures of the semantic relationships among menu set stimuli. The specific procedures and results of this work are detailed in Section III of this report.

Synopsis

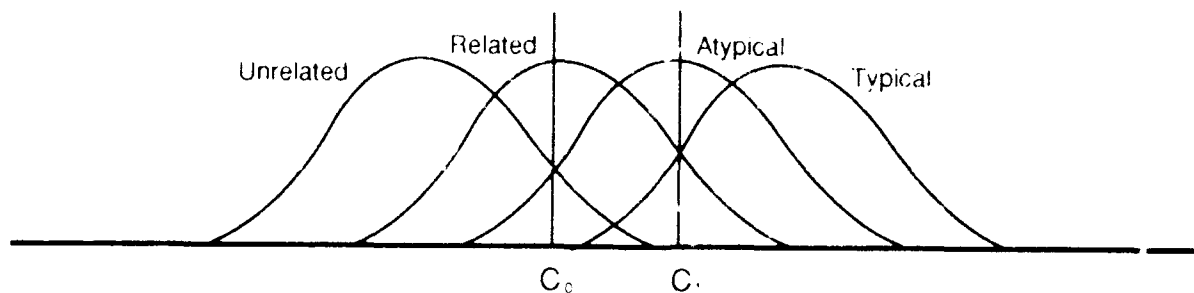
The material reviewed in this section suggests that the relationships among targets and menu alternatives in a computer menu selection task, such as the one used in the study by MacGregor et al. (1986), can be conceptualized in terms of semantic distances among menu set items derived from MDS solutions of similarity ratings. In the present investigation, these relationships were defined and measured operationally as two distinct measures. The first measure reflected the relationship between the target and the correct alternative in the menu display ("semantic distance to correct" or D_c). This measure was operationally defined as the semantic distance between these items derived from MDS solutions of similarity ratings. The second measure reflected the amount of competition for selection produced by the incorrect alternatives presented in the menu display ("semantic distance to incorrect" or D_i). The D_i measure was defined operationally in more than one manner. The competition produced by incorrect alternatives was defined as either a distance measure from correct to incorrect alternatives or a distance measure from the target to incorrect alternatives. Note that as either of these distance measures increases, the selection competition produced by incorrect alternatives in the menu decreases. In addition, each measure was operationally defined either as a measure of the central tendency among distance values for all competing alternatives or as a single measure of distance to the semantically nearest neighbor in the menu display.

It was decided to attempt to derive distance measures among menu set items that would allow for the computation of all these measurement variants. Resolution as to which measure best accounts for performance variations could then be determined empirically. Section III details

the results of this work. For the moment, D_i is defined as a measure of semantic distance that is inversely proportional to the level of selection competition produced by incorrect alternatives. Thus, as D_i decreases, competition for selection increases.

From the review of semantic memory research, changes in D_c and D_i are predicted to affect both menu selection search time and accuracy. Specifically, the lower the value of D_c (reflecting a high degree of similarity between target and correct alternative), the lower will be both the processing time and the probability of error resulting from the evaluation of a correct alternative. Conversely, the lower the value of D_i (reflecting a high degree of selection competition from incorrect alternatives), the greater will be the processing time and the probability of error resulting from the evaluation of individual incorrect alternatives.

The effects that semantic similarities among targets and alternatives have on selection speed and accuracy can be conceptualized in terms of variations in the distributions of relatedness among targets and alternatives with respect to selection criteria. Consider, for example, the two-stage semantic memory model proposed by Smith, Shoben, and Rips (1974). Figure 8 provides a schematic representation of the hypothesized effects that variations in relatedness distributions have on response processes during a category verification task. The figure illustrates the process by which a subject determines whether test instances are members of a particular category. The two distributions on the right side represent probability distributions for instances for typical and atypical category members. The two other distributions represent distributions for related and unrelated nonmembers. According to Smith et al.'s (1974) model, the subject performs this verification task by first comparing the instance to the category and then by determining the overall semantic similarity, x , of the category to the instance. C_0 and C_1 are selection criteria used to determine whether to execute a fast true response ($x > C_1$) or a fast false response ($x < C_0$), or to go on to a second comparison stage ($C_0 < x < C_1$). Smith et al. (1974) tested predicted performances of this model and showed that, to a large extent, results supported the conceptualization of semantic similarity effects in the manner specified.



Semantic Relatedness -->

Figure 8 Hypothetical Distribution of Overall Similarity Values Between Targets and Category Names (adapted from Smith et al., 1974).

The effects of semantic similarities among menu set items in a menu selection task can be conceptualized in a similar manner by generalizing from the single instance/category verification problem. In the menu selection task where the subject attempts to determine under which of the menu alternatives the target is subsumed, the axis in Figure 8 would no longer represent the relatedness of the target to a single category. Rather, the axis would represent a more general function of the degree of semantic relatedness that can be used to illustrate the relationship between the target and any alternative considered for selection. In addition, rather

than representing probabilities for members and nonmembers of a single category, the distributions would represent probabilities for members of the correct alternative (illustrated by the two distributions to the right) and the overall probability for nonmembers of all incorrect alternatives (illustrated by the distributions to the left). Assuming that D_c and D_i values have been scaled to represent common metrics of semantic relatedness, the axis in Figure 8 can be thought of as representing these same metrics. (Note that both D_c and D_i decrease as semantic similarity increases, or as values move left to right along the axis.)

The probability distribution representing the semantic similarity between correct alternatives and typical targets is illustrated in Figure 8 by the curve farthest to the right. The mean D_c value associated with this distribution would be relatively low, reflecting a high degree of semantic relatedness. The probability distribution representing the relationship between correct alternatives and atypical targets is illustrated by the second curve from the right. The mean D_c value for this distribution would be somewhat higher than that for the distribution labeled "Typical," reflecting decreased semantic relatedness. The probability distributions of semantic relatedness of targets to related and unrelated incorrect alternatives are represented by the two curves on the left. The mean D_i values for these distributions would be relatively higher than the mean D_c values associated with the other two distributions, reflecting decreased relatedness. Ignoring for the moment the placement of selection criteria, it is apparent from this illustration that variations in the probability distributions of D_c and D_i values essentially amount to variations in the discriminability among correct and incorrect alternatives.

In terms of Smith et al.'s (1974) model of semantic memory, such variations have been shown to affect processing times of individual alternatives, as well as response accuracy, in category verification tasks. In terms of the menu search model proposed by MacGregor et al. (1986), variations in D_c and D_i distributions might be expected to affect menu search strategies as well. To the extent that such search strategy variations exist, as was evidenced by the results of MacGregor et al.'s (1986) experiment, a model analogous to the one proposed by Smith et al. (1974) is proposed that predicts such search strategy variations as being partially due to variations in the discriminability among menu alternatives. Details of this model are presented in Section V and are used in Section VI to explicate the results of the menu task experiment described in the next section.

Research Rationale

The background literature reviewed in this section suggests that several factors potentially contribute to the creation of an optimal menu design: (a) the number of alternatives displayed on the menu, (b) the semantic relationships among targets and menu item alternatives, and (c) the probability of a correct alternative omission. Though the literature has shown these factors to be predictive of performance on menu or menu-related tasks, the effects these factors have on menu task performance have never been studied under conditions where all factors are manipulated within a single experimental design. Such an experiment was the focus of the present investigation. The design permitted the assessment of effects each factor had on task performance. In addition, through the independent estimation of menu item semantic relationships, potential confounding due to experimental manipulation of omission probability and number of alternatives was also evaluated.

The first step was to select a menu set appropriate for the current investigation and to quantify semantic relationships among menu item members. Menu set selection was based on requirements that the menu represent a command-type menu hierarchy with a categorical structure that could be manipulated logically for experimental purposes. An additional requirement was that semantic relationships among the menu items within the set had to be quantified. This requirement stemmed from the need to verify that experimental manipulations made to menu stimuli actually resulted in variations of semantic relationships among menu targets and alternatives.

Miller's (1981) menu hierarchy (Figure 1) fulfilled the requirements for a command-type menu, with appropriate hierarchical relationships among its members. In addition, an extensive performance database has been developed for this menu through various experiments reported in the literature and could be used to help validate predictions and results of the present investigation. Although the concept and measurement of the number of menu alternatives and omission probability are relatively straightforward, measurement of semantic relationships among targets and menu alternatives is more complex. The requirement to quantify the relationships among the items contained in the menu set was satisfied through the collection of similarity ratings from groups of subjects for pairwise combinations of menu stimuli. Distance values were estimated from the rating data using MDS procedures. Upon completion of the scaling work, the menu set then was used in an experiment designed to quantify the performance effects of the four factors of interest (Dc, Di, omission probability and number of alternatives). Details of the scaling and experimental methods are described in the next section. The results of this work are explicated by means of a two-criterion menu model which is detailed in Section V of this report.

III. METHODS

Purpose

The purpose of the present experiment was to quantify the performance effects of four menu design factors on subject performances. The four factors were: (a) the hierarchical level of the menu page (i.e., the level within the menu hierarchy from which the menu was derived); (b) the hierarchical relationship or "nesting" that existed among menu set alternatives (i.e., the level in the hierarchy at which all menu alternatives had a common superordinate); (c) the number of alternatives that were presented on the menu; and (d) the probability that the correct alternative was omitted from the menu. Hereafter, these factors will be referred to as the Hierarchy, Nesting, Number of Alternatives, and Omission Probability factors, respectively. A final factor used in this experiment was sequential versus simultaneous presentation of menu stimuli. Hereafter, this factor will be referred to as the Presentation Mode factor. The sequential condition provided direct measures of search strategies. Most command menu implementations, however, involve simultaneous presentation of alternatives. Use of both conditions allowed comparison of response accuracy data obtained for each condition.

Determining Semantic Relationships Among Menu Stimuli

The task stimuli used in this investigation represented a command-type menu hierarchy with a categorical structure that could be manipulated logically for experimental purposes. Additionally, semantic relationships among the stimulus items within the set had to be quantified. This second requirement stemmed from the need to verify that experimental manipulations made to menu stimuli actually resulted in variations of semantic relationships among menu targets and alternatives.

Miller's (1981) menu hierarchy (Figure 1) met the requirements for a command-type menu with appropriate hierarchical relationships. In addition, an extensive performance data base had been developed using this menu that could be used for comparison to predictions and results of the present investigation. The requirement to quantify the relationships among the items contained in this set was satisfied through a scaling procedure in which pairwise ratings of semantic similarity for menu set items were collected from groups of subjects and then scaled using multidimensional scaling (MDS) techniques. A summary of the general approach and results of these scaling efforts is described below.¹

¹A detailed summary of the semantic scaling effort appears in Pierce (1989).

Menu Set Selection

The design for the menu task experiment required the selection and scaling of only a subset of the items in Miller's (1981) original hierarchy (see Figure 1). The newly formed menu set contained all 64 targets from Level VI, the 32 items from Level V, and the eight items from Level III. Items contained in the resulting menu set are shown in Figure 9.

All items from Level VI of Miller's (1981) hierarchy were used as menu targets. Targets are those category items labeled 1 through 8 in Figure 9. The eight items taken from Level III of Miller's (1981) hierarchy were used to construct menus with alternatives having a distant hierarchical relationship to Level VI targets. Level III alternatives are listed as item 13 of each category set in Figure 9. The 32 items taken from Level V of the hierarchy were used to construct menus with alternatives having a close hierarchical relationship to Level VI targets. Level V alternatives are numbered 9 through 12 in each category listed in Figure 9.

Scaling Requirements

An assessment of the impact that menu item relationships had on task performances required a measure of the semantic relatedness between the targets and the correct menu alternatives (measures of D_c), and measures reflecting the amount of selection competition produced by incorrect menu alternatives (measures of D_i). A series of data collection procedures were designed such that these scaling requirements could be met in one of two ways.

The first approach was to develop measures of the distances among all items that formed the menu set. From these measures, distances from targets to all alternatives would be available for computing target-to-correct-alternative similarity measures, as well as a variety of measures reflecting the semantic relatedness between targets and incorrect menu alternatives. Traditional approaches to multidimensional scaling (MDS) of semantic similarities among members of a word set require ratings for all possible pairwise combinations of every member of the set. The rating data then are scaled to a best-fit multidimensional configuration using one of several statistical programs designed for this purpose (see Schiffman, Reynolds, & Young, 1981).

To obtain rating data for all possible pairwise combinations of the 104 items in the current menu set would require 5,356 comparisons per subject. An alternative scaling approach was examined which produced a set of distance values for all possible pairwise combinations of menu items from a much smaller subset of pairwise similarity ratings (Homa & Konrad, 1987; Konrad, 1988). The technique was used to "conjoin" MDS solutions of pairwise similarity data collected for subsets of the 104-item menu set. Each subset had several items in common with other subsets to which it was conjoined. The conjoining algorithm took the two subsets of scaled elements, defined by n -dimensional coordinates, and used the common item coordinates of one subset to scale, translate, and orient those coordinates into maximal congruence with the coordinates of the other subset.

There were three main operations involved in producing the conjoined solution. First, the algorithm equated the extent or volume of the common elements, resulting in an expansion/contraction of one of the two scaled spaces. Second, the common elements were translated such that, in the final solution, they were centered around the same point in space. Finally, the two spaces were conjoined such that the pairwise Euclidean distances of the common elements were minimized. Although this technique greatly reduced the number of pairwise comparisons required of individual subjects to produce its solution, the technique was new and the validity of its results had yet to be tested fully. For this reason, data collection procedures were designed to provide not only measures of inter-item distances among menu set members, but also a means for testing the validity of the resulting solutions.

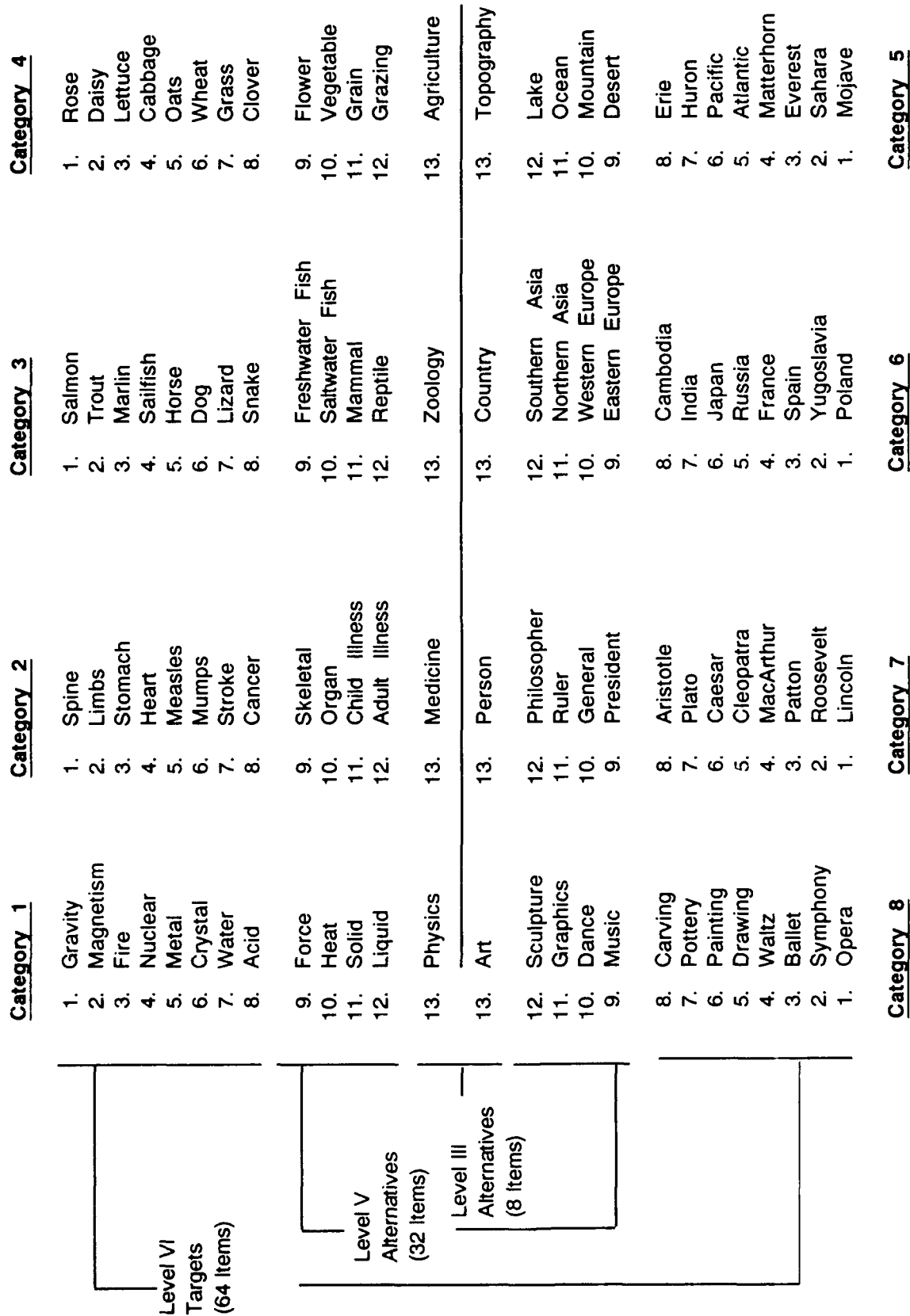


Figure 9. Menu Set Items Derived from Levels III, V, and VI of Miller's (1981) Hierarchy.

Scaling procedures also were designed such that, should the validity of the conjoining procedures prove questionable, a second approach to meeting scaling requirements could be implemented without the need for additional rating data. This approach used the MDS solutions for the menu item subsets without employing the conjoining procedure. The approach produced estimates of semantic distances among items within category sets, as well as estimates of the semantic distances among the 40 Levels III and V alternatives used to construct menu pages for the menu task experiment. Thus, distances from targets to correct alternatives within menus were estimated under this approach, as were distances from the correct menu page alternative to incorrect ones. The main drawback of this approach was that it did not produce distance estimates from the target to alternatives outside the target's immediate category set. Because, as will be detailed later in this section, some of the menu pages were constructed using alternatives from outside the target's own category, this scaling approach did not allow the computation of distances from targets to incorrect alternatives contained in these menu pages. Measures of selection competition produced by incorrect alternatives that could be computed from this second approach, therefore, were limited to include only estimates of the distances from the correct menu alternative to incorrect ones.

Results Summary

Measures of congruence between those solutions which used the conjoin procedure and those solutions which did not indicated that the conjoin solutions were of questionable validity. Therefore, it was decided to use distance measures derived directly from MDS solutions of subject rating averages for separate sets of pairwise combinations of menu items. Similarity ratings were obtained for all possible combinations of the 40 alternatives from Levels III and V (Figure 9). In addition, separate rating sets were obtained for each grouping of category set items. Thus, the 13-item category sets also depicted in Figure 9 formed eight groups of pairwise combinations which were individually rated for similarity by subject groups.

Individual MDS solutions were derived for these nine rating sets (the set of 40 alternatives from Levels III and V and the eight 13-item category sets). Resulting solutions for each of the eight 13-item category sets then were scaled to a set of interpoint distances (IPDs) derived from items common to the set of 40 alternatives from Levels III and V. The resulting IPD values for the MDS solution of the set of 40 alternatives provided semantic similarity estimates among correct and incorrect alternatives used to construct menu set displays for the menu task experiment. These measures were used to compute values of the semantic distance between correct and incorrect menu alternatives (the D_i distance measure). IPD values derived from the scaled MDS solutions for the eight 13-item category sets were used to compute semantic similarity estimates among targets and correct alternatives (the D_c distance values). The resulting MDS solutions for the individual data sets had low stress values, and correlations between common-item IPDs from independently derived data sets were relatively high. These results indicated that the goodness of fit between the distance estimates derived from these solutions and the original similarity rating averages was good to excellent (according to goodness-of-fit classifications by Kruskal, 1964), and that the distance values were highly reliable in terms of replication. Based on these findings, it was concluded that the validity of the distance measures derived from these solutions was highly satisfactory.

Although distance data derived from MDS solutions of individual data sets were deemed acceptable in terms of validity, they did not provide distances from targets to cross-category alternatives. It should be reiterated that the original objective of the scaling work was to derive distance measures among all possible pairwise combinations of menu stimuli so that an optimal measure for D_i could be empirically determined. Without distance measures between targets and alternatives from other categories, the operational definition of the selection competition measure (D_i) was limited to include only estimates of the distances from the correct menu

alternative to incorrect ones. Should the results of this experiment confirm hypothesized performance effects, the impact of the limited Di measure would simply be the problem of generalizing the results to broader definitions of menu set discriminability produced by competing alternatives. On the other hand, should the results show no significant effects for the Di factor, it would not be possible to distinguish whether these results were due to the true absence of a main effect or, rather, due to an inadequate level of measurement sensitivity/reliability associated with the limited Di measure.

In summary, even without distance estimates for targets to cross-category items, the scaling procedures did achieve the objective of providing a set of semantic proximity measures to be used to quantify menu item relationships in the menu task experiment. These data provided the means to evaluate semantic manipulations of menus and the effect these manipulations had on task performances.

General Approach

A partial search task procedure was used. Subjects attempted to match a target, presented at the onset of each trial, with the correct alternative contained in a 2-, 4- or 8-alternative menu, which was displayed following target presentation. Following examination of the menu display, subjects made a response from options available for that menu page. If the response resulted in an error, the subject continued with that trial until the correct response was made. Both simultaneous and sequential modes of presenting alternatives were used. In the simultaneous mode, all alternatives were displayed vertically at the same time. In the sequential mode, alternatives were displayed one at a time in a sequential, top-down manner. Rate of presentation in this latter condition was controlled by the subject. The simultaneous presentation mode provided data that were used only to determine accuracy of subject responses. The sequential presentation condition provided these measures, as well as a direct measure of search strategy for individual trial responses. The combination of sequential and simultaneous procedures allowed the study of decision and search processes employed by the subjects, as well as permitting the assessment of whether the forced sequential search imposed on subjects in the sequential condition influenced response accuracy.

Experimental Design

A mixed experimental design was used, with two between-groups factors and three within-group factors (see Table 4). The first between-groups factor was the Presentation Mode--either sequential or simultaneous--used for display of alternatives. The second between-groups factor, the Omission Probability factor, represented four levels of omission trial probability (i.e., 0%, 12.5%, 25%, and 37.5%). The first within-group factor was the Number of Alternatives factor, representing the number of alternatives presented on the menu page (either 2, 4, or 8). The second within factor was the Hierarchy factor, representing the two levels in the menu hierarchy from which each menu was derived. These levels were defined by the hierarchical distance between the target and the correct alternative in the menu. In one level of the Hierarchy factor, menus were constructed using alternatives from Level III of the Figure 1 hierarchy. In the other level, menus were constructed using alternatives from Level V. The third within factor was two levels of the Nesting factor, representing that level in the hierarchy which contained a superordinate common to all alternatives in the menu display. Thus, these levels were defined by the hierarchical distance between correct and incorrect alternatives in the menu. In the Nesting factor close condition, menus were constructed such that the relationship between the correct alternative and all incorrect alternatives in the menu was as hierarchically close as possible. In the Nesting factor distant condition, menus were constructed such that the hierarchical relationship between correct and incorrect alternatives was relatively distant. The Nesting factor was manipulated only within the 4-alternative menu condition. For the 2- and 8-alternative menus, the Nesting factor was fixed at the close level.

Table 4. Experimental Design

Between-groups factors			Within-group factors							
(12 subjects per group)			Hierarchy Level III				Hierarchy Level V			
			Number of alternatives				Number of alternatives			
			2	4	4	8	2	4	4	8
Group	Presentation mode	Omission probability	Nesting				Nesting			
			close	close	dist	close	close	close	dist	close
1	Seq	0%								
2		12.5%								
3		25%								
4		37.5%								
5	Sim	0%								
6		12.5%								
7		25%								
8		37.5%								

Note. Seq - sequential presentation mode; Sim - simultaneous presentation mode.

Apparatus

IBM PC-compatible computers with 12-inch green monochrome monitors were employed to present stimuli, and record responses.

Subjects

The subjects were undergraduate college students enrolled in an introductory psychology course at Arizona State University. The design required eight groups, with twelve subjects per group.

Procedure

Each subject performed two practice trials, followed by 128 experimental trials. The 2-alternative menu trials, 4-alternative Nesting distant trials, 4-alternative Nesting-close trials, and 8-alternative trials were all grouped across the two Hierarchy factor levels. This resulted in four trial sets, with 32 trials per set. Within-group ordering effects were controlled by counterbalancing the sequencing of trial sets across subjects. All subjects within each group performed identical sets of trials using the same target-to-menu-condition pairings, positioning of menu alternatives within trials, and omission trial assignments. Presentation order of trial sets associated with the within-group factors was counterbalanced across subjects. Order of trial presentation within trial sets was randomized across the first group of subjects. Subjects in both the sequential and simultaneous conditions, as well as the Omission Probability conditions, were then yoked with respect to trial presentation order. Thus, the only procedural difference between subject groups was the stimulus Presentation Mode crossed by Omission Probability; the only procedural difference among subjects within groups was the presentation order of trials.

Each trial set began with presentation of a message which described the number of trials to be completed (32 for each set) and the Number of Alternatives contained in each menu. Individual trials began with presentation of the target. Subjects initiated presentation of alternatives by pressing the space bar.

Procedures similar to those employed by MacGregor et al. (1986) were employed for both the sequential and simultaneous Presentation Mode conditions. In the sequential condition, only the identification numbers corresponding to each alternative were initially visible. Subjects displayed the first alternative in the menu by striking the space bar. Subjects controlled cursor movement on the screen by pressing the "up" and "down" special function keys. Subjects displayed alternatives by moving the cursor to its corresponding position on the screen and striking the space bar. Any previously displayed alternative then was erased from the screen and the newly selected alternative was displayed in its own position. Subjects could display the next alternative in the menu sequence or any previously viewed alternatives; however, they could not skip ahead in the menu over alternatives not viewed. In the simultaneous condition, all alternatives were displayed on the screen at the same time. In all conditions, the alternatives on each page were arranged in a column, with one alternative per line.

In both simultaneous and sequential conditions, subjects made their choice by pressing the "A" key, followed by the number of the alternative selected (or zero if the subject believed that the target was not to be found under any of the alternatives on the page). In the sequential condition, only the alternatives viewed were available for selection. The zero response was allowed only after all menu alternatives were viewed. Thus, guessing was not permitted. A correct response in either condition led to presentation of the next search target. An incorrect or unacceptable response led to an error message and a request to try again. In the sequential condition, unacceptable responses indicative of guessing were not treated as errors but rather, as an inappropriate key response. In this situation, a message would be displayed which informed the subject of the problem and requested that the subject continue the trial. Subjects continued each trial until the correct response was made. At any time, a subject could redisplay the target by pressing the "T" key.

Materials

Menu Construction

Assumptions. The semantically scaled menu set items were used in the current experiment to form the 45 menu pages depicted in Tables 5 and 6. Manipulations of the semantic relationships among targets and menu alternatives were based on two assumptions derived from semantic memory models described in the literature (e.g., Collins & Quillian, 1969; Rips et al., 1973). The first assumption was that the semantic distance between a target and a correct alternative is directly related to the hierarchical distance between them. Thus, as the hierarchical distance increases, the semantic distance will correspondingly increase. Based on this assumption, the distances between targets and correct alternatives (the Hierarchy factor) were manipulated by pairing targets (Level VI items in Figure 1) with alternatives from either Level III or Level V of the Figure 1 hierarchy.

The second assumption on which semantic distance manipulations were based was that menu alternatives sharing a common superordinate at the same level in the menu hierarchy were semantically closer to each other than to alternatives sharing common origins at higher levels in the hierarchy. For example, the categories "Flower," "Vegetable," "Grain," and "Grazing" all stem from the same Level III superordinate "Agriculture" in Miller's (1981) hierarchy (see Figure 1). "Flower" and "Vegetable," however, also share the hierarchically closer Level IV item "Garden." Because neither "Grain" nor "Grazing" is subordinate to "Garden," the assumption

argues that "Flower" and "Vegetable" should be closer semantically to one another than either item is to "Grain" or "Grazing." Based on this assumption, the distances among the menu alternatives (the Nesting factor) were manipulated by grouping alternatives based on their superordinate origins at various levels in Miller's (1981) original hierarchy. With the exception of menus contained in the Nesting distant condition (see Table 4), all menu alternatives were grouped as hierarchically close as possible. It should be noted that these latter menus were arranged as they would have appeared had the user actually navigated to that menu level from superordinate levels in the hierarchy.

Table 5. Menus Constructed from Level III of the Hierarchy (9 menus total)

2-Alternative Menu (4 menus)

Menu #1	Menu #2	Menu #3	Menu #4
Agriculture	Medicine	Country	Art
Physics	Zoology	Topography	Person

4-Alternative Menus

Nesting Close (2 menus)		Nesting Distant (2 menus)	
Menu #1	Menu #2	Menu #1	Menu #2
Agriculture	Country	Agriculture	Physics
Physics	Topography	Medicine	Zoology
Medicine	Art	Country	Topography
Zoology	Person	Art	Person

8-Alternative Menu (1 menu)

Agriculture
Physics
Medicine
Zoology
Country
Topography
Art
Person

Menu Page Preparation. Based on the assumptions and procedures outlined above, menus for trials in which a distant semantic relationship was desired between the target and correct alternative were prepared using Level III alternatives in the hierarchy depicted in Figure 1. Menus for trials in which a close semantic relationship was desired between the target and correct alternative contained only Level V alternatives. Nine menu pages using Level III alternatives were generated (see Table 5): Four pages contained two closely related alternatives; two pages contained four closely related alternatives; two pages contained four distantly related alternatives; and one page contained eight closely related alternatives. Thirty-six menu pages were developed using alternatives selected from Level V of the hierarchy (review Table 6): Sixteen pages contained two closely related alternatives; eight pages contained four closely related alternatives; eight pages contained four distantly related alternatives; and four pages contained eight closely related alternatives.

To minimize the semantic distance among the four alternatives used in menus constructed for the Hierarchy-Level III x Nesting-close x 4-Alternatives condition, alternatives from Level III of the hierarchy stemming from a common superordinate in Level I of Figure 1 were grouped together. Menus constructed for the Hierarchy-Level V x Nesting-close x 4-Alternatives condition were designed by grouping alternatives from Level V of the hierarchy which had a common superordinate at Level III of the hierarchy. To maximize the semantic distance among the 4

Table 6. Menus Constructed from Level V of the Hierarchy (36 menus total)

2-Alternative Menus (16 menus)

Menu #1	Menu #2	Menu #3	Menu #4
Flower	Grain	Force	Solid
Vegetable	Grazing	Heat	Liquid
Menu #5	Menu #6	Menu #7	Menu #8
Skeletal	Child Illness	Freshwater Fish	Mammal
Organ	Adult Illness	Saltwater Fish	Reptile
Menu #9	Menu #10	Menu #11	Menu #12
Eastern Europe	Northern Asia	Desert	Ocean
Western Europe	Southern Asia	Mountain	Lake
Menu #13	Menu #14	Menu #15	Menu #16
Music	Graphics	President	Ruler
Dance	Sculpture	General	Philosopher

4-Alternative Menus

Nesting Close (8 menus)

Menu #1	Menu #2	Menu #3	Menu #4
Flower	Freshwater Fish	Skeletal	Force
Vegetable	Saltwater Fish	Organ	Heat
Grain	Mammal	Child Illness	Solid
Grazing	Reptile	Adult Illness	Liquid
Menu #5	Menu #6	Menu #7	Menu #8
Eastern Europe	Desert	Music	President
Western Europe	Mountain	Dance	General
Northern Asia	Ocean	Graphics	Ruler
Southern Asia	Lake	Sculpture	Philosopher

Nesting Distant (8 menus)

Menu #1	Menu #2	Menu #3	Menu #4
Flower	Vegetable	Grain	Solid
Skeletal	Organ	Child Illness	Mammal
Eastern Europe	Western Europe	Northern Asia	Ocean
Music	Dance	Graphics	Ruler
Menu #5	Menu #6	Menu #7	Menu #8
Force	Heat	Grazing	Liquid
Freshwater Fish	Saltwater Fish	Adult Illness	Reptile
Desert	Mountain	Southern Asia	Lake
President	General	Sculpture	Philosopher

Table 6 (Concluded)

8-Alternative Menus (4 menus)			
Menu #1	Menu #2	Menu #3	Menu #4
Music	Eastern Europe	Skeletal	Flower
Dance	Western Europe	Organ	Vegetable
Graphics	Northern Asia	Child Illness	Grain
Sculpture	Southern Asia	Adult Illness	Grazing
President	Desert	Freshwater	Fish Force
General	Mountain	Saltwater	Fish Heat
Ruler	Ocean	Mammal	Solid
Philosopher	Lake	Reptile	Liquid

alternatives used in the Hierarchy-Level III x Nesting-distant x 4-Alternatives condition, alternatives from Level III in the hierarchy stemming from separate superordinates at Level II of the hierarchy were used to form each menu. Similarly, menus in the Hierarchy-Level V x Nesting-distant x 4-Alternatives condition were created by selecting alternatives from Level V in the hierarchy having common Level I origins yet stemming from separate superordinates at Level III of the hierarchy.

Target-to-Menu Pairings

The 64 Level VI targets were divided into 16 groups consisting of four targets each such that all targets in each group shared a common Level IV superordinate (see Figure 1). Targets in each group were then assigned for use with each of the four conditions defined by levels of the Nesting and Alternatives factors. This procedure was done separately for each of the two levels of the Hierarchy factor. For each of the 16 target groups, assignment of targets to menu conditions was made using a random selection technique (with the restriction that only one target from each group could be assigned to each of the four menu conditions). Once target-to-menu-condition assignments had been completed, the menu page containing the correct alternative for the target assigned to that condition was then paired with the target stimulus. Within each of the two levels of the Hierarchy factor, individual targets were used only once. Each target therefore was used twice across both levels of this factor. As a result of this procedure, each subject performed a total of 128 trials. Ordering of alternatives was randomized for each menu display (with the restriction that the location of the correct alternative on each menu display was distributed equally across trials over the available positions).

Verifying Assumptions

Problem Issues

The intent of the Hierarchy manipulation was to produce two levels of menu stimuli which differed significantly with respect to the semantic relationship between the target and the correct alternative. Similarly, the intent of the Nesting manipulation was to produce two levels of menu stimuli which significantly differed with respect to the semantic relationship between correct and incorrect alternatives. The intent of the Alternatives manipulation was simply to create menus in which the Number of Alternatives presented for selection varied from level to level. Although the intent of the Alternatives manipulation was readily verified by counting Number of Alternatives, the variation of semantic relationships resulting for the manipulations associated with the Hierarchy and Nesting factors required a more sophisticated examination.

Another issue requiring resolution concerned the potential confounding effects of manipulations on semantic relationships among stimuli. Specifically, the potential effects of variations in the Alternatives and Hierarchy factors on the semantic relationships between correct and incorrect alternatives needed to be assessed. Similarly, the potential effects of variations in the Alternatives and Nesting factors on the semantic relationships between targets and correct alternatives used in the menus needed to be examined.

Analytic Procedures

Semantic proximity manipulations for the menus and trial arrangements defined above were tested statistically through a series of ANOVAs using as dependent variables the distance values estimated from the previously described scaling procedures. Two sets of ANOVAs were computed. In the first set, the effects due to manipulations of the Alternatives and Hierarchy factors were analyzed using the following three dependent measures in separate analyses: (a) scaled distances between targets and correct menu alternative (scaled distance to correct alternative - D_c), (b) scaled distance between the correct alternative and the semantically nearest incorrect alternative in the menu (scaled distance to nearest neighbor - D_i -NN), and (c) the average of the scaled distances between the correct alternative and all incorrect alternatives in the menu (D_i -AVE). The second set of ANOVAs tested the effects that manipulations of the Nesting and Hierarchy factors had on the same dependent variables examined in the first ANOVA set. These procedures not only allowed the assessment of the intent of the menu manipulations, but also allowed the assessment of potential confounding among the design factors with respect to D_c and D_i measures. What follows is a summary of the procedures, findings and implications of these ANOVAs.²

Findings and Implications

The analyses of manipulation effects on scaled distance values have several implications concerning the effects of the factor manipulations and the type of analytic procedures that should be used when interpreting menu task performance. First, findings involving the Hierarchy and Nesting factors were consistent with respect to the expected effects of these factors on the semantic relationships among targets and alternatives. Level III menus consistently had significantly greater mean D_c values than did Level V menus. Second, the distant level of the Nesting factor consistently had significantly greater mean D_i -NN and D_i -AVE values than did the close level.

Of equal importance was the finding that the main effects on mean D_i -NN and D_i -AVE measures for both the Alternatives and the Hierarchy factors were consistently significant in the first set of ANOVAs. In addition, the Hierarchy by Nesting interaction on mean D_i measures was consistently significant in the second set of ANOVAs. These findings suggest a confounding of both the Alternatives and the Hierarchy factors due to the effect these factors have on the semantic relationship between correct and incorrect alternatives.

The confounding among these factors implied that in order to assess the effects of these variables on menu task performance, a statistical technique was required that would allow the removal of unwanted sources of variance for each variable analyzed. Regression analysis provided the capability to partial out shared variance for each predictor with respect to all other

²A detailed summary of the procedures and findings of the menu manipulation ANOVAs appears in Pierce (1989).

predictors contained in the model. Significance testing of the partial coefficients then permitted the required assessment of the unique amount of variance accounted for by each variable in the regression model. The Dc, Di-NN and Di-AVE continuous measures for each trial were used in place of the levels associated with the Hierarchy and Nesting factors. Using these variables, a regression analysis of performance data which removed unwanted variance was conducted. Section IV gives a complete description of this procedure.

Omission Trials

Omission trial Probability was established in each of the 2- and 4-alternative menu sets by replacing the correct alternatives with incorrect ones. The number of omission trials presented in a given block of trials was set at the between-groups condition rate for each group of subjects (Omission Probability factor in Table 4). The four probability levels used were typical of error rates obtained in previous menu navigation studies (e.g.; MacGregor et al., 1986; Snowberry et al., 1983). Omission trials for the 8-alternative menus were arranged somewhat differently. Because the 8-alternative menus in the Hierarchy Level III condition contained all possible alternatives from that level, only one menu was constructed for the Hierarchy Level III x 8 Alternatives condition (see Table 5). Thus, replacement of the correct Level III alternative with an incorrect one from the menu set was not possible. Therefore, all Omission trials in the 8-alternative trials were constructed using Hierarchy Level V menus.

Construction of Omission trials proceeded as follows. Semantic distances between the correct menu alternative and its semantically closest neighbor that was not a member of the menu in question were calculated for each trial. For each Hierarchy Level III menu, only Level III alternatives were considered to be neighbors. In a similar manner, only Level V alternatives were considered for Hierarchy Level V menus. Within each trial condition, trials having the closest distances between the correct alternative and its nearest available neighbor were selected to be Omission trials. For those trials so selected, the correct alternative was replaced with its nearest available neighbor. In trial sets where the same correct alternative was used more than one time with the same menu, once selected the alternative was not considered again until all other available alternatives had been used an equal number of times.

The above procedures were intended to provide Omission trial settings that were not highly obvious to the subject. The use of these procedures, however, may have introduced two potential sources of confounding effects. First, by selecting the semantically nearest available incorrect alternative as a replacement, Di-NN and Di-AVE values may have decreased with increasing levels of the Omission factor. The second potential confounding source concerns the effect the Omission trial selection procedure had on Dc values. Omission trials were formed using those trials having the closest distances between the correct alternative and its nearest available neighbor; therefore, to the extent that variations in Dc and Di-NN measures may have been correlated, the possibility exists that mean Dc values for trials in which the Correct Alternative was Present were also confounded across levels of the Omission Probability factor.

The effect that Omission trial selection procedures had on Dc, Di-NN and Di-AVE values was examined analytically by computing two sets of ANOVAs, using an approach similar to the one used to verify menu manipulations. Two sets, each consisting of three ANOVAs, were computed. In the first set, the effects on each of three semantic distance measures separately were tested for manipulations of the Alternatives, Hierarchy, and Omission factors, along with a fourth factor which identified whether or not the Correct Alternative was Present in the menu. The Nesting factor was fixed at the close level in this first set of analyses. In the second set of ANOVAs, the manipulations of the Hierarchy, Nesting, Omission and Correct Alternative Present factors were tested in three separate ANOVAs for their effects on Dc, Di-NN and Di-AVE. Thus, Dc, Di-NN and Di-AVE values for individual trials were the basic unit of analysis in both sets of ANOVAs. In all ANOVAs, the Dc distance measures for Correct Alternative Absent trials represented the distance from the target to the correct alternative that was removed from the menu to form the Omission trial.

None of the tests in either set of analyses indicated significant main effects for the Omission factor. In addition, no interaction terms containing the Omission factor had significant effects on any of the dependent variables. However, the main effect for the factor representing the Presence or Absence of the Correct Alternative was significant in all ANOVAs, as were several interaction terms containing this factor. Examination of Table 7 indicates that means for all three distance measures were greater for the Correct Alternative Present trials than they were for Correct Alternative Absent trials. The results suggest that the Omission trial selection procedure did not have a significant effect on the relationships among targets and alternatives across levels of the Omission factor. However, the results also suggest that a stronger relationship exists between targets and incorrect menu alternatives when the Correct Alternative is Absent than when the Correct Alternative is Present.

Table 7. Comparison of Mean Dc, Di-NN and Di-AVE Values for Correct Alternative Present versus Absent Trials

	Correct alternative	
	Present	Absent
Fixed: Nesting close		
Dc <u>M</u>	.759	.626
<u>n</u>	(312)	(72)
Di-NN <u>M</u>	.983	.693
<u>n</u>	(312)	(72)
Di-AVE <u>M</u>	1.21	1.02
<u>n</u>	(312)	(72)
Fixed: Alternatives = 4		
Dc <u>M</u>	.782	.676
<u>n</u>	(208)	(48)
Di-NN <u>M</u>	1.143	.714
<u>n</u>	(208)	(48)
Di-AVE <u>M</u>	1.402	1.162
<u>n</u>	(208)	(48)

Note. Upper means for all trials where Nesting was fixed at close; lower means for all trials where Alternatives were fixed at 4.

The results of these analyses suggest that separate analyses of Correct Alternative Present data and Correct Alternative Absent data are required. These results, combined with the confounding of the Alternatives and Hierarchy factors noted earlier from analysis of menu stimuli manipulations, formed the basis for the regression procedures used during the analysis of menu task performance data.

IV. RESULTS

The performance data were analyzed in two phases. The first phase consisted of a series of regression analyses to test the effects of Alternatives, Omission Probability, Dc values and Di values on response accuracy. An additional factor examined was Presentation Mode (i.e., sequential versus simultaneous). The second phase focused on the effects of Alternatives,

Omission Probability, Dc values and Di values on subjects' search strategies using sequential Presentation Mode data only.

Phase I - Response Accuracy

Procedures

Response accuracy data analysis procedures followed from the implications given by the menu manipulation and Omission trial selection analyses described in Section III. The results of these analyses suggested that performance data could best be interpreted using a regression approach. Such an approach would allow the analysis of the Alternatives, Omission Probability, Dc, and Di effects by computing the partial coefficients for each of these predictors and then testing these coefficients for significance. The results of the Omission trial selection analyses also suggested that performances on Correct Alternative Present trials should be analyzed separately from Correct Alternative Absent trials. Consequently, data associated with individual trials were separated into two data sets as defined by the Presence or Absence of a Correct Alternative. Each data set was analyzed separately.

Correct Alternative Present Trials. Proportions of hit responses for Correct Alternative Present trials are summarized in Table 8 for individual cells of the experimental design. Regression analysis of hit responses proceeded as follows. Proportions of hit responses for individual trials were computed across subjects for each of the eight groups and used as the basic unit of analysis. Following procedures recommended by Cohen and Cohen (1983) for transformation of proportion data, the arcsine transformations of these proportions were used as the criterion variable. Separate regression analyses were computed for the sequential Presentation Mode groups, the simultaneous Mode groups, and the combination of data from both groups. For each data set, two regression analyses were computed. One used the Di-NN value and the other used the Di-AVE value as the predictor variable to define the relationship between correct and incorrect alternatives. In all regression analyses, additional predictors included the Number of Alternatives, Omission Probability and the Dc trial value. For regression analyses in which data from both Presentation Modes were combined, a predictor variable representing Presentation Mode was added to the above.

Correct Alternative Absent Trials

Proportions of correct rejections for Correct Alternative Absent trials are summarized in Table 9 for individual cells. Procedures similar to those above were used for regression analysis of correct rejection data. Proportions of correct rejection responses were computed across subjects within each group for individual trials and used as the basic unit of analysis. The arcsine transformations of these proportions were computed and used as the criterion variable. Because all trials in the 0% Omission Probability condition contained the correct alternative, correct rejection performance data from only six groups of subjects were used. As with the Correct Alternative Present analyses, separate analyses were performed on data from each Presentation Mode in addition to the data from both modes combined. Predictor variables were identical to those used in the analysis of Correct Alternative Present response accuracy data.

Findings

Correct Alternative Present Trials

Table 10 summarizes the regression analyses results for Correct Alternative Present trial performances. The F-values for both the Di-NN and the Di-AVE predictor models were significant

Table 8. Cell-by-Cell Proportions of Hits for Correct Alternative Present Trials

Group	Presentation mode	Omission probability	Hierarchy Level III				Hierarchy Level V				Marginal Proportions (across all menus)			
			Number of alternatives				Number of alternatives							
			2	4	4	8	2	4	4	8				
1	Sequential	0%	Nesting				Nesting							
			close	close	dist	close	close	close	dist	close				
			.990	.958	.896	.849	.870	.844	.979	.813				
			(192)	(192)	(192)	(192)	(192)	(192)	(192)	(192)				
2	Sequential	12.5%	.798	.833	.821	.781	.893	.851	.863	.743	.824			
			(168)	(168)	(168)	(192)	(168)	(168)	(168)	(144)	(1344)			
			.806	.763	.778	.786	.951	.819	.889	.677	.813			
			(144)	(144)	(144)	(192)	(144)	(144)	(144)	(96)	(1152)			
4	Sequential	37.5%	.750	.750	.708	.672	.967	.875	.875	.771	.789			
			(120)	(120)	(120)	(192)	(120)	(120)	(120)	(48)	(960)			
			1-4	Sequential	All	.849	.840	.813	.772	.913	.846	.907	.760	.838
						(624)	(624)	(624)	(768)	(624)	(624)	(624)	(480)	(4992)
5	Simultaneous	0%				.922	.943	.885	.891	.865	.818	.932	.745	.875
						(92)	(192)	(192)	(192)	(192)	(192)	(192)	(192)	(1536)
			6	Simultaneous	12.5%	.893	.857	.875	.818	.958	.851	.917	.757	.867
						(168)	(168)	(168)	(192)	(168)	(168)	(168)	(144)	(1344)
7	Simultaneous	25%				.819	.784	.882	.828	.924	.868	.903	.729	.846
						(144)	(144)	(144)	(192)	(144)	(144)	(144)	(96)	(1152)
			8	Simultaneous	37.5%	.733	.750	.775	.719	.925	.850	.892	.917	.805
						(120)	(120)	(120)	(192)	(120)	(12)	(120)	(48)	(960)
5-8	Simultaneous	All				.854	.846	.861	.814	.915	.845	.913	.763	.853
						(624)	(624)	(624)	(768)	(624)	(624)	(624)	(480)	(4992)
			Marginal Proportions (across all groups)											
			.852	.843	.837	.793	.914	.845	.910	.761	.845			
			(1248)	(1248)	(1248)	(1536)	(1248)	(1248)	(1248)	(960)	(9984)			
Note. Number of observations in parentheses.														

Note. Number of observations in parentheses

Table 9. Cell-by-Cell Proportions of Correct Rejections for Correct Alternative Absent Trials

Group	Presentation mode	Omission probability	Hierarchy Level III				Hierarchy Level V				Marginal Proportions (across all menus)
			Number of alternatives				Number of alternatives				
			2	4	4	8	2	4	4	8	
			close	close	dist	close	close	close	dist	close	
2	Sequential	12.5%	.708 (24)	.750 (24)	1.000 (24)	.	.042 (24)	.458 (24)	.375 (24)	.688 (48)	.589 (192)
3	Sequential	25%	.792 (48)	.583 (48)	.979 (48)	.	.229 (48)	.375 (48)	.438 (48)	.510 (96)	.552 (384)
4	Sequential	37.5%	.861 (72)	.847 (72)	.917 (72)	.	.417 (72)	.639 (72)	.708 (72)	.597 (144)	.698 (576)
2-4	Sequential	All	.813 (144)	.743 (144)	.951 (144)	.	.292 (144)	.521 (144)	.563 (144)	.583 (288)	.631 (1152)
6	Simultaneous	12.5%	.667 (24)	.750 (24)	1.000 (24)	.	.167 (24)	.458 (24)	.417 (24)	.750 (48)	.620 (192)
7	Simultaneous	25%	.792 (48)	.708 (48)	1.000 (48)	.	.333 (48)	.438 (48)	.646 (48)	.563 (96)	.630 (384)
8	Simultaneous	37.5%	.889 (72)	.819 (72)	.792 (72)	.	.444 (72)	.486 (72)	.667 (72)	.625 (144)	.668 (576)
6-8	Simultaneous	All	.819 (144)	.771 (144)	.896 (144)	.	.361 (144)	.465 (144)	.618 (144)	.625 (288)	.648 (1152)
Marginal Proportions (across all groups)			.816 (288)	.757 (288)	.924 (288)	.	.326 (288)	.493 (288)	.590 (288)	.604 (576)	.639 (2304)
Note. Number of observations in parentheses.											

Note. Number of observations in parentheses.

Table 10. Regression Results for Proportion of Hits on Correct Alternative Present Trials

ANOVA Results												
Sequential Mode				Simultaneous Mode				Both Modes Combined				
4 Predictor Models				4 Predictor Models				5 Predictor Models				
Source	DF	SS	MS	F	DF	SS	MS	F	DF	SS	MS	F
Di - NN Model (nearest neighbor)	4	41.078	10.270	34.161***	4	27.508	6.877	23.787***	5	66.618	13.324	45.018***
Error	411	123.554	.301		411	118.821	.289		826	244.466	.296	
Total	415	164.633			415	146.329			831	311.084		
Di - AVE Model (average distance)	4	41.736	10.434	34.894***	4	29.247	7.312	25.667***	5	68.917	13.783	47.013***
Error	411	122.897	.299		411	117.082	.285		826	242.167	.293	
Total	415	164.633			415	146.329		831	831	311.084		
Predictor Variables												
	pr	Coef	Std Err	t	pr	Coef	Std Err	t	pr	Coef	Std Err	t
A. Di - NN Model												
Intercept		3.191	.115	27.753***		2.872	.113	25.472***		3.043	.083	36.738***
Alternatives	.017	-.035	.013	-2.703**	.011	-.027	.013	-2.106*	.014	-.031	.009	-3.398***
Omission Prob	.052	-.928	.196	-4.735***	.014	-.473	.192	-2.459*	.030	-.700	.138	-5.093***
Dc Distance	.171	-1.015	.110	-9.199***	.131	-.850	.109	-7.860***	.149	-.933	.077	-12.049***
Di NN Distance	.048	.378	.083	4.554***	.077	.476	.081	5.837***	.061	.427	.058	7.325***
Presentation Mode			N/A				N/A		.000	-.024	.038	-.641
B. Di - AVE Model												
Intercept		3.203	.112	28.705***		2.876	.109	26.409***		3.052	.080	37.980***
Alternatives	.056	-.062	.012	-4.938***	.057	-.061	.012	-4.965***	.056	-.061	.009	-6.987***
Omission Prob	.050	-.90	.195	-4.647***	.013	-.448	.191	-2.350*	.029	-.678	.137	-4.956***
Dc Distance	.175	-1.042	.112	-9.342***	.141	-.895	.109	-8.226***	.157	-.969	.078	-12.405***
Di AVE Distance	.053	.403	.084	4.801***	.090	.523	.082	6.378***	.070	.463	.059	7.874***
Presentation Mode			N/A				N/A		.001	-.024	.038	-.644
Cumulative R-squares												
Alternatives												
+ Omission Prob		.048		.043		.043		.045				
+ Dc Distance		.095		.056		.056		.073				
+ Di Distance		.212		.121		.121		.163				
+ Presentation Mode	NN - .250	.254		NN - .189	.200	NN - .214	.221	NN - .2141	.2215			
	N/A			N/A		N/A		N/A				

*p < .05.

01
V
010
★★

L/D
V
.001.

for the sequential Mode data, the simultaneous Mode data, and the combined Modes data. R-square values show the cumulative proportion of variance accounted for with the addition of the predictor variables listed. The cumulative R-squares for the addition of the Di-NN and Di-AVE predictors indicate that both variables account for approximately the same amount of variance for all three data sets analyzed.

Partial correlations and coefficients for all of the model predictor variables, with the exception of the coefficient for the Presentation Mode variable, were statistically significant in all analyses. The tests of the partial correlations and coefficients indicate the significance of the unique variance accounted for by the Alternatives, Omission Probability, Dc, and Di predictors.

Comparison of the partial correlations and coefficients derived for the sequential Presentation Mode data to the same statistics derived for the simultaneous Mode data indicates a high degree of similarity. Partial coefficients resulting from the regressions for these two data sets all indicate the same positive or negative relationship to the criterion; and, for the most part, the partial correlations were of about the same magnitude. One exception was the partial correlation derived for the Omission Probability predictor. Though coefficients for this variable were negative in all regressions, the Omission Probability partial correlations for sequential Mode were more than triple the values obtained for the simultaneous Mode. This finding would seem to indicate that Omission Probability had a more pronounced effect on the criterion in the sequential Presentation Mode than in the simultaneous Presentation Mode.

The correlations and coefficients for the Presentation Mode predictor were not significant in either regression where both Modes were combined. Additional analyses of two-way interaction terms which included the Presentation Mode variable indicated that none of the other variables in the model significantly interacted with Presentation Mode ($p > .05$). Thus, the results suggest that Presentation Mode did not have a significant main effect, nor did it significantly moderate the effects of other predictor variables. These results are consistent with the hypothesis that similar decision-making processes are used in both Modes and that the Mode factor does not have a significant effect on decision accuracy.

Examination of the significant partial coefficients for model predictors indicates the following relationships were consistent across all three data sets: (a) As the Number of Alternatives, Omission Probability or semantic distance between target and correct alternative increased, proportion of hits decreased; (b) conversely, as the semantic distance (as defined by either Di-NN or Di-AVE) between correct and incorrect alternatives increased, the proportion of hits increased. These results are consistent with the hypothesis that response accuracy on Correct Alternative Present trials is significantly affected by variations in these four menu design factors.

Correct Alternative Absent Trials

Results of the regression analyses for Correct Alternative Absent data are presented in Table 11. F-values for the regressions were significant in each of the six analyses computed for these data. R-square values show the cumulative proportion of variance accounted for with the addition of the predictor variables listed. In contrast to the cumulative R-squares obtained from analysis of Correct Alternative Present data, the cumulative R-squares obtained for the present analyses indicate that the Di-NN predictor accounted for an additional 5% to 11% of the variance beyond that accounted for by the Di-AVE predictor across the three data sets.

Examination of the R-square values, along with the results of the significance tests for the partial correlations and coefficients, indicates that the variance accounted for by these models was due mainly to the inclusion of the Di predictors. In the Di-NN models, the only significant partial correlations and coefficients were for the Di-NN predictor across all three data sets. As

Table 11. Regression Results for Proportion of Correct Rejections
on Correct Alternative Absent Trials

ANOVA Results		Sequential Mode				Simultaneous Mode				Both Modes Combined			
		4 Predictor Models				4 Predictor Models				5 Predictor Models			
A. Di - NN Model (nearest neighbor)		DF	SS	MS	F	DF	SS	MS	F	DF	SS	MS	F
Source		4	16.825	4.206	10.370***	4	12.950	3.238	7.446***	5	29.634	5.927	14.350***
Model		91	36.912	.406		91	39.569	.435		186	76.821	.413	
Error		95	53.737			95	52.519			191	106.455		
Total													
B. Di - AVE Model (average distance)		4	11.062	2.766	5.897***	4	10.406	2.602	5.621***	5	21.338	4.268	9.326***
Source		91	42.674	.469		91	42.113	.463		186	85.117	.458	
Model		95	53.737			95	52.519			191	106.455		
Error													
Total													
Predictor Variables		pr	Coef	Std Err	t	pr	Coef	Std Err	t	pr	Coef	Std Err	t
A. Di - NN Model													
Intercept		.000	.439	.340	1.292		.861	.352	2.447*		.682	.247	2.764***
Alternatives		.009	.002	.030	.066	.000	.003	.031	.110	.000	.003	.022	.126
Omission Prob		.005	.625	.706	.885	.000	-.062	.731	-.085	.002	.281	.504	.558
Dc Distance		.266	-.231	.341	-.676	.009	-.322	.353	-.912	.007	-.276	.243	-1.136
Di NN Distance			2.074	.361	5.737***	.224	1.920	.374	5.130***	.244	1.997	.258	7.743***
Presentation Mode				N/A				N/A		.003	-.064	.093	-.694
B. Di - AVE Model													
Intercept		.052	.691	.360	1.918		1.008	.358	2.819**		.882	.256	3.441***
Alternatives		.008	-.084	.037	-2.240*	.058	-.087	.037	-2.361*	.003	-.085	.026	-3.283***
Omission Prob		.003	.655	.761	.860	.000	-.086	.756	-.114	.013	.284	.531	.535
Dc Distance		.151	.189	.347	.544	.000	-.013	.345	-.038	.038	.088	.242	.363
Di AVE Distance			1.192	.296	4.023***	.174	1.291	.294	4.385***	.162	1.241	.207	5.998***
Presentation Mode				N/A				N/A		.002	-.064	.098	-.659
Cumulative R-squares													
Alternatives			.003				.002					.002	
+ Omission Prob			.025				.005					.012	
+ Dc Distance			.065				.029					.044	
+ Di Distance			NN - .313	AVE - .206			NN - .247	AVE - .198			NN - .277	AVE - .199	
+ Presentation Mode			N/A				N/A				NN - .278	AVE - .200	

*p < .05.
**p < .01.
***p < .001.

Di-NN values increased, proportions of correct rejections increased. In the Di-AVE models, partial correlations and coefficients for both the Di-AVE predictor and the Alternatives predictor were significant. As Di-AVE values increased, proportions of correct rejections increased. Conversely, as the Number of Alternatives increased, proportions of correct rejections decreased.

As with the analysis of the Correct Alternative Present response accuracy data, the comparison of significant partial correlations and coefficients across Presentation Modes indicates a high degree of similarity. Significant coefficients all have the same sign and partial correlations all have similar values. The partial correlations and coefficients derived for the Presentation Mode predictor were not significant in either analysis where data from Presentation Modes were combined. Further testing of two-way interaction terms which included the Presentation Mode variable indicated that none of the other predictors in the model significantly interacted with Presentation Mode ($p > .05$). These results complement those obtained from analysis of the Correct Alternative Present data, suggesting that the Presentation Mode did not have a significant main effect on response accuracy, nor did it significantly moderate the effects of other model predictors.

Performance Consistency Across Presentation Modes

Regression results indicated a high degree of consistency between subject performances obtained in the two Presentation Modes. This consistency is further exemplified in Tables 12 and 13. These tables show correlations, means, and standard deviations of response accuracy measures for sequential and simultaneous Mode data sets within cells of trials as defined by the Alternatives and Omission factors. Table 12 contains these values for arcsine transformations of the proportion of hits for Correct Alternative Present trials. Table 13 contains values for arcsine transformations of the proportion of correct rejections for Correct Alternative Absent trials. The overall correlation between Presentation Modes for arcsine transformations of the proportions of hits was .674 ($p < .001$). For arcsine transformations of the proportions of correct rejections, the overall correlation was .855 ($p < .001$). The correlation results complement the findings from the response accuracy analyses, suggesting that subjects' response performances from the two Presentation Modes were comparable.

Phase II - Search Strategies

Phase II analyzed the effects of performance factors on the search strategies employed in the sequential Presentation condition. With the data generated from this experiment, there was no direct means to verify the generalization of search strategy findings from analysis of sequential mode data to simultaneous presentation of alternatives. However, the preceding analyses of response accuracy data are consistent with the hypothesis that such a generalization is appropriate.

Procedures

Search strategy regression analyses proceeded as follows. For every trial performed in the sequential Mode condition, the numbers of different and redundant alternatives examined prior to making a response were recorded. Search strategy was inferred from these data and classified as self-terminating, exhaustive or redundant using the following operational definitions. Self-terminating searches were defined as those wherein the subject selected an alternative prior to examination of all alternatives in the menu. Exhaustive search was defined as one in which all alternatives in the menu were examined once and only once prior to selection. Redundant search was defined as a search in which all menu alternatives were examined and some alternatives re-examined prior to selection. These definitions imply that in situations where the

subject examines all alternatives once and only once and then chooses the alternative occupying the last position of the menu, search strategy would be classified as exhaustive. It could be argued, however, that such searches could be classified as self-terminating in that the subject terminated the search upon examination of the alternative selected. Therefore, search strategy for these situations was considered indeterminate and its use in analyses was subject to the procedures detailed below.

Table 12. Correlations Between Presentation Modes for Arcsine Transformations of Proportions of Hits--Correct Alternative Present Trials Only

Omission probability	Number of alternatives		
	2	4	8
0%	r = .644*** n = 32	r = .734*** n = 64	r = .495** n = 32
	M seq = 2.851	M seq = 2.821	M seq = 2.457
	SD seq = .518	SD seq = .575	SD seq = .576
	M sim = 2.666	M sim = 2.670	M sim = 2.365
	SD sim = .511	SD sim = .523	SD sim = .512
12.5%	r = .512** n = 28	r = .757*** n = 56	r = .684*** n = 28
	M seq = 2.486	M seq = 2.513	M seq = 2.255
	SD seq = .524	SD seq = .614	SD seq = .629
	M sim = 2.777	M sim = 2.618	M sim = 2.335
	SD sim = .452	SD sim = .568	SD sim = .722
25%	r = .619** n = 24	r = .659*** n = 48	r = .688*** n = 24
	M seq = 2.611	M seq = 2.393	M seq = 2.206
	SD seq = .525	SD seq = .614	SD seq = .608
	M sim = 2.565	M sim = 2.524	M sim = 2.344
	SD sim = .521	SD sim = .522	SD sim = .610
37.5%	r = .678*** n = 20	r = .684*** n = 40	r = .662** n = 20
	M seq = 2.574	M seq = 2.389	M seq = 2.104
	SD seq = .613	SD seq = .640	SD seq = .788
	M sim = 2.515	M sim = 2.466	M sim = 2.199
	SD sim = .714	SD sim = .692	SD sim = .741
Overall:	r = .674*** n = 416		
	M seq = 2.508	M sim = 2.533	
	SD seq = .630	SD sim = .594	

**p < .01.

***p < .001.

Table 13. Correlations Between Presentation Modes for Arcsine Transformations of Proportions of Correct Rejections – Correct Alternative Absent Trials Only

Omission probability	Number of alternatives					
	2		4		8	
	r = .982*	n = 4	r = .983***	n = 8	r = .911	n = 4
12.5%	M seq = 1.283		M seq = 2.071		M seq = 1.959	
	SD seq = 1.366		SD seq = .953		SD seq = .171	
	M sim = 1.410		M sim = 2.083		M sim = 2.106	
	SD sim = .808		SD sim = 1.007		SD sim = .225	
	r = .958***	n = 8	r = .870***	n = 16	r = .664	n = 8
25%	M seq = 1.639		M seq = 1.870		M seq = 1.592	
	SD seq = .939		SD seq = .807		SD seq = .410	
	M sim = 1.821		M sim = 2.156		M sim = 1.689	
	SD sim = .998		SD sim = .801		SD sim = .544	
	r = .890***	n = 12	r = .795***	n = 24	r = .817**	n = 12
37.5%	M seq = 1.967		M seq = 2.246		M seq = 1.779	
	SD seq = .764		SD seq = .486		SD seq = .772	
	M sim = 2.095		M sim = 2.037		M sim = 1.917	
	SD sim = .909		SD sim = .576		SD sim = .660	
Overall:	r = .855***	n = 96				
	M seq = 1.918		M sim = 1.983			
	SD seq = .752		SD sim = .746			

*p < .05.

**p < .01.

***p < .001.

Three sets of regression analyses were computed. Because subjects were limited to employing only exhaustive and redundant searches on trials where the Correct Alternative was Absent (assuming a correct response), only trials in which the Correct Alternative was Present were used in these analyses. In the first set of regressions, arcsine transformations of the proportion of self-terminating searches were computed across each group of 12 subjects for individual trials. Only trials for which all 12 subjects exhibited determinable search strategies were included in the analyses. Two regressions were performed in this set. The first included the Di-NN variable as the predictor representing the relationship between correct and incorrect alternatives. The second included the Di-AVE variable as the predictor representing this relationship. In both regressions, Alternatives, Omission Probability and Dc values were also included as model predictors with the arcsine transforms of the proportions of self-terminating searches serving as the criterion.

In the second set of regressions, arcsine transformations of the proportion of exhaustive searches for individual trials were used as the criterion. Proportions of exhaustive searches were computed across each group of 12 subjects for individual trials. Following the same procedure used for self-terminating search data, only Correct Alternative Present trials for which all 12 subjects exhibited determinable search strategies were included in the analyses. Two

regressions were computed, one using Di-NN as the predictor representing the relationship between the correct and incorrect alternatives, and the other using Di-AVE as the predictor. In both regressions, additional model predictors were Alternatives, Omission Probability, and Dc values.

In the third regression set, arcsine transformations of the proportion of redundant searches for individual trials were used as the criterion. Proportions of redundant searches were computed across each group of 12 subjects for individual trials. Because computation of proportions of redundant searches were not confounded by indeterminable search situations, data from all Correct Alternative Present trials were included in these analyses. In addition, to be consistent with analyses of self-terminating and exhaustive search data, regression analyses of redundant searches were also computed for Correct Alternative Present trials for which all 12 subjects exhibited determinable search strategies. As with the previously described search strategy regression procedures, two regressions were computed for each data set, using Di-NN as a model predictor in one and the Di-AVE variable in its place in the other. In all regressions, Alternatives, Omission Probability and Dc values were included as additional model predictors.

Findings

Proportions of self-terminating, exhaustive and redundant searches for Correct Alternative Present trials for which all 12 subjects exhibited determinable search strategies are summarized in Table 14 for each cell included under the sequential Mode condition. Regression results obtained for the redundant search data indicated a high degree of consistency between analyses of data for all Correct Alternative Present trials versus trials on which all 12 subjects exhibited determinable search strategies. To be consistent with self-terminating and exhaustive search analyses, only the results of the latter analyses of redundant searches are presented. Table 15 lists the results of the six search strategy regressions computed for trials where all 12 subjects exhibited determinable searches. F-tests for the regression models were significant in all six analyses.

Examination of the partial correlations and coefficients, and significance test results for the self-terminating search model predictors shows that all predictors, with the exception of the Di predictors, significantly contributed to the model regressions. As either the Number of Alternatives or Omission Probability increased, self-terminating searches increased. As the semantic distance between targets and correct alternatives increased, self-terminating searches decreased. Cumulative R-square values listed at the bottom of Table 15 indicate that both the Di-NN and the Di-AVE models accounted for approximately 37% of the self-terminating search variance.

The results from analysis of exhaustive search data indicate that only the Alternatives and Dc predictors significantly contributed to the model regressions. Examination of the partial coefficients for these variables indicated that as the Number of Alternatives increased, exhaustive searches increased. Cumulative R-square values indicate that approximately 36% of the variance was accounted for by both the Di-NN and the Di-AVE models.

All four partial correlations and coefficients in the redundant search models significantly contributed to the regressions in both the Di-NN and Di-AVE models. Examination of the coefficients indicates that as the Number of Alternatives and as the semantic distance between target and the correct alternative increased, the proportion of redundant searches increased. Redundant searches decreased as Omission Probability and the semantic distance from correct to incorrect alternatives increased. Approximately 28% of the variance was accounted for by both the Di-NN and Di-AVE regression models.

In all three sets of regressions for the search strategy data, the comparison of the Di-NN versus Di-AVE regression models indicated little difference between them in terms of significance

Table 14. Cell-by-Cell Proportions of Self-Terminating, Exhaustive and Redundant Searches for Correct Alternative Present Trials Where All Subjects Exhibited Determinable Search Strategies

Group	Omission probability	Search type	Hierarchy Level III					Hierarchy Level V					Marginal Proportions (across all menus)
			Number of Alternatives					Number of Alternatives					
			2	4	8	close	dist	close	dist	close	dist	close	
			Nesting					Nesting					
1	0%	Self-Terminating Exhaustive Redundant	.181 .708 .111 (72)	.146 .604 .250 (144)	.265 .614 .121 (132)	.312 .382 .306 (144)	.450 .517 .033 (60)	.417 .483 .100 (120)	.269 .635 .096 (156)	.468 .308 .224 (156)	.311 .518 .171 (984)		
2	12.5%	Self-Terminating Exhaustive Redundant	.146 .771 .083 (48)	.130 .805 .065 (108)	.183 .750 .067 (120)	.264 .604 .132 (144)	.444 .556 .000 (36)	.317 .633 .050 (120)	.317 .675 .008 (120)	.408 .517 .075 (120)	.272 .662 .066 (816)		
3	25%	Self-Terminating Exhaustive Redundant	.056 .944 .000 (36)	.104 .802 .094 (96)	.274 .678 .048 (84)	.404 .449 .147 (156)	.403 .597 .000 (72)	.381 .571 .048 (84)	.519 .435 .046 (108)	.625 .278 .097 (72)	.367 .559 .074 (708)		
4	37.5%	Self-Terminating Exhaustive Redundant	.055 .917 .028 (36)	.178 .655 .167 (84)	.306 .625 .069 (72)	.378 .494 .128 (156)	.458 .500 .042 (48)	.393 .476 .131 (84)	.479 .469 .052 (96)	.625 .292 .083 (24)	.357 .543 .100 (600)		
Marginal Proportions (across all groups)		Self-Terminating Exhaustive Redundant	.125 .807 .068 (192)	.139 .708 .153 (432)	.250 .669 .081 (408)	.342 .482 .176 (600)	.435 .546 .019 (216)	.375 .544 .081 (408)	.379 .567 .054 (480)	.489 .368 .142 (372)	.322 .570 .108 (3,108)		
Note. Number of observations in parentheses.													

Note. Number of observations in parentheses.

Table 15. Regression Results for Search Type Proportions
on Correct Alternative Present Trials

ANOVA Results									
A. Di - NN Model (nearest neighbor)									
Source	DF	SS	MS	F	DF	SS	MS	F	DF
Model	4	31.314	7.828	37.993***	4	23.188	5.797	36.198***	4
Error	254	52.337	.206		254	40.677	.160		254
Total	258	83.650			258	63.865			258
B. Di - AVE Model (average distance)									
Source	DF	SS	MS	F	DF	SS	MS	F	DF
Model	4	31.373	7.843	38.109***	4	23.110	5.777	36.007***	4
Error	254	52.277	.206		254	40.755	.160		254
Total	258	83.650			258	63.865			258
Predictor Variables									
	pr	Coef	Std Err	t	pr	Coef	Std Err	t	
A. Di - NN Model									
Intercept		1.649	.130	12.662***		1.564	.115	13.619***	.103
Alternatives	.076	.063	.014	4.565***	.199	-.097	.012	-7.938***	.011
Omission Prob	.028	.551	.205	2.696**	.000	-.059	.180	-.328	.162
Dc Distance	.273	-1.204	.123	-9.760***	.150	.729	.109	6.706***	.098
Di NN Distance	.000	-.003	.095	-.033	.006	.102	.084	1.209	.076
B. Di - AVE Model									
Intercept		1.607	.131	12.265***		1.577	.116	13.628***	.103
Alternatives	.081	.063	.013	4.729***	.232	-.103	.012	-8.763***	.010
Omission Prob	.028	.554	.205	2.710**	.000	-.052	.181	-.289	.161
Dc Distance	.283	-1.241	.124	-10.019***	.153	.741	.109	6.772***	.097
Di AVE Distance	.001	.052	.096	.539	.004	.083	.084	.986	.075
Cumulative R-Squares									
Alternatives		.044				.173			
+ Omission Prob		.050				.174			
+ Dc Distance		.374				.359			
+ Di Distance		.375				.363			
	NN	.374	AVE	.375	NN	.363	AVE	.363	NN
									.290
									AVE
									.302

*p < .05.
**p < .01.
***p < .001.

of their partial correlations and coefficients or in terms of variance accounted for by the models. It is interesting to consider the relative effects of the Alternatives, Omission Probability, Dc, and Di variables with respect to the three search strategy types as a whole. As the Number of Alternatives increased in the menu, the proportion of both self-terminating and redundant searches tended to increase at the expense of exhaustive searches. As Omission Probability increased, the proportion of self-terminating searches increased at the expense of redundant searches, although this variable had no significant effect on exhaustive searches. Increases in the distance between the target and correct alternative increased the probability of both exhaustive and redundant searches at the expense of decreases in self-terminating searches. Lastly, as the distance between the correct alternative and incorrect alternatives increased, redundant searches decreased, but the effect of this variable with respect to either self-terminating or exhaustive searches was not significant. The overall effects demonstrated through these analyses for variations in the Alternatives, Omission Probability, Dc and Di variables support the hypothesis that these factors significantly influence search strategy performance on computer menu tasks.

Synopsis

Analysis of search strategy and response accuracy data indicated the significant effects that Alternatives, Omission Probability, and Dc and Di semantic distance measures had on menu task performance. Furthermore, analyses of response accuracy data for sequential and simultaneous Presentation Modes suggested that subjects' decision performances from the two Modes were comparable. This latter result supports the generalization of search strategy findings from the sequential Mode to simultaneous presentation of alternatives.

Although the analyses reported here established the significance of menu design factors with respect to task performance, their implications with respect to cognitive processes involved in menu search and decision making are limited. The next section provides an overview of a two-criterion menu model which suggests a process that can potentially account for the effects of menu design factors on task performance. The next section also details the results of a data-fitting procedure designed to assess the relationship between model criteria and menu design factors. Further discussion of the results of the menu task performance analyses and implications for menu design are then presented in Section VI in the context of the two-criterion menu model.

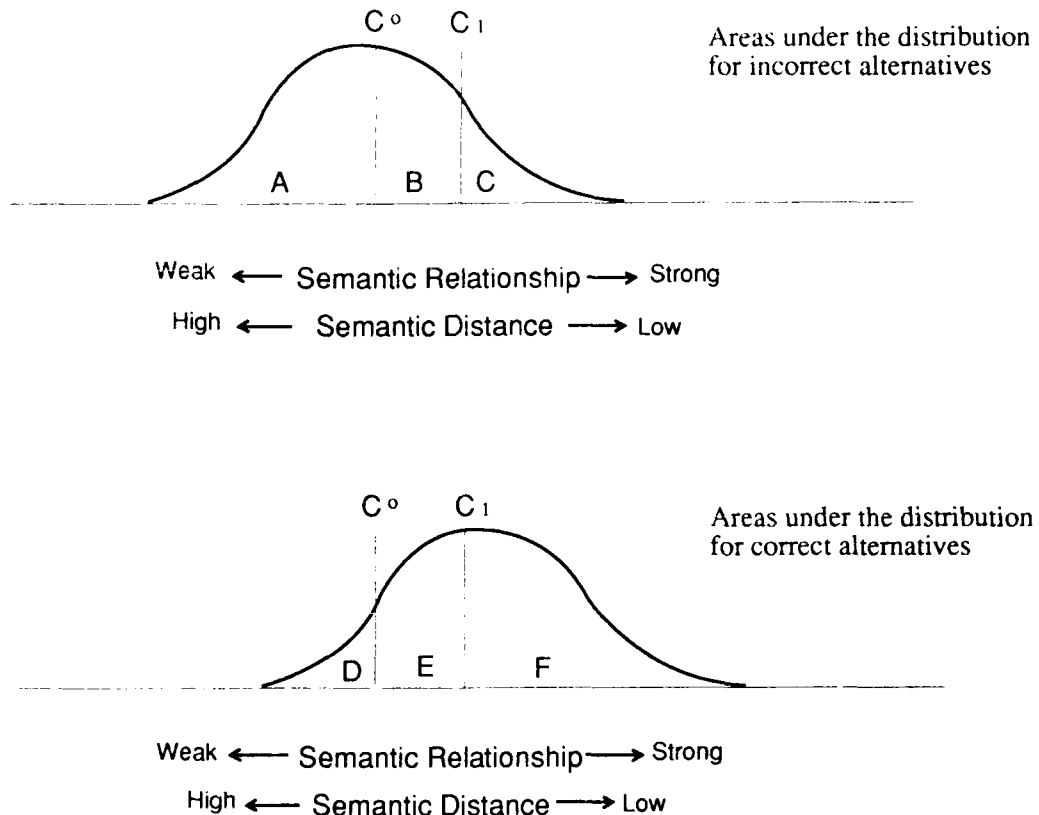
V. THE TWO-CRITERION MENU MODEL

Overview

The concept of a two criterion menu model that could be used to predict menu search strategies and response outcomes was initially proposed in Section II. The genesis of this concept was the Smith et al. (1974) feature comparison model that was developed to explain cognitive processes involved in semantic verification tasks. The two-criterion menu model extends concepts proposed by Smith et al. (1974) to the multiple verification problem involved in menu selection tasks. This section elaborates the development of the two-criterion menu model and details procedures used to fit performance data obtained from the menu task experiment to model prediction functions. Essentially, performance data were fitted by deriving values for model parameters that minimized differences between model predictions and obtained performances. The derived values for model parameters were then analyzed through a series of regressions to determine the relationship between model criteria and menu design factors. The results of these procedures were used to explicate the cognitive processes involved during performance of the menu search task.

Description

The two-criterion menu model was designed to predict combinations of search strategy and response outcome probabilities under a variety of experimental conditions for the partial search menu task. The model assumes that subjects evaluate alternatives in the menu display in a sequential fashion and determine the overall semantic similarity, x , of each alternative examined with respect to the target. The distribution of relatedness values among targets and incorrect alternatives for a given menu set is depicted by the upper curve in Figure 10. The lower curve in this figure represents the distribution of relatedness values between targets and correct alternatives for the same menu set. Both curves in this figure are Gaussian.



Legend.

C0 - Criterion for immediate rejection. All alternatives whose relatedness value falls to the left of this criterion are predicted to be immediately rejected.

C1 - Criterion for immediate selection. All alternatives whose relatedness value falls to the right of this criterion are predicted to be immediately selected (thus leading to self-terminating search).

A through F - Areas defined under the probability distributions by the two criteria.

Figure 10. Areas Defined Under the Distribution Curves for the Two-Criterion Menu Model.

The criterion C0 in Figure 10 represents the criterion for immediate rejection. Any alternative whose semantic relatedness to the target (x) falls below this value ($x < C0$) will be immediately rejected as a response candidate. The menu search for the correct alternative will therefore

continue. C1 represents the criterion for immediate selection of an alternative without further evaluation of remaining alternatives. Thus, evaluation of such alternatives is predicted to result in a self-terminating search ($x > C1$). Evaluation of alternatives with relatedness values between these criteria results in the continued search of additional alternatives. If none of the remaining alternatives searched have relatedness values that exceed C1, the alternative possessing the highest degree of similarity to the target will be selected, provided its relatedness value exceeds C0. Exhaustive searches result when all but one of the alternatives are rejected without further consideration and the similarity between the target and the remaining alternative is sufficient to continue search but insufficient to trigger self-termination. Exhaustive searches are also predicted for situations where all alternatives are rejected ($x < C0$), resulting in a zero response. When none of the alternatives exceeds the rapid selection criterion C1 and similarity values for more than one of the alternatives fall between criteria, all of these latter alternatives are re-evaluated, resulting in a predicted redundant search.

The probability that a correct or an incorrect alternative falls within a specific range of semantic relatedness values is identified by areas A through F in Figure 10. These areas are functions of the parameters which characterize the two Gaussian distributions depicted in this figure and the location of selection criteria. Formulas for computing these areas are given in Figure 11. These areas can be used to estimate the probability that a single alternative from a given menu set has a relatedness value that is below, in between, or above criteria values. In addition, when values for Number of Alternatives and Omission Probability are known, probability functions associated with areas A through F can be used to estimate the probabilities of combinations of search strategies (self-terminating, exhaustive and redundant) and response outcomes (hit, commission miss, false alarm, omission miss, and correct rejection) for a given menu set.

$$\text{Area A} = \int_{-\infty}^{C0} e^{-\frac{1}{2} [x - \mu_{Xi}]^2} dx$$

$$\text{Area B} = \int_{C0}^{C1} e^{-\frac{1}{2} [x - \mu_{Xi}]^2} dx$$

$$\text{Area C} = \int_{C1}^{\infty} e^{-\frac{1}{2} [x - \mu_{Xi}]^2} dx$$

$$\text{Area D} = \int_{-\infty}^{C0} e^{-\frac{1}{2} [x - \mu_{Xc}]^2} dx$$

$$\text{Area E} = \int_{C0}^{C1} e^{-\frac{1}{2} [x - \mu_{Xc}]^2} dx$$

$$\text{Area F} = \int_{C1}^{\infty} e^{-\frac{1}{2} [x - \mu_{Xc}]^2} dx$$

Legend.

C0 - Rapid rejection criterion value.

C1 - Rapid selection criterion value.

μ_{Xc} - Mean relatedness value for correct alternative distribution.

μ_{Xi} - Mean relatedness value for incorrect alternative distribution.

Figure 11. Equations for Computing Areas Under Probability Distributions for the Two-Criterion Menu Model.

Eleven combinations of search strategy and response outcomes were predicted by the model. For self-terminating and redundant searches, the model predicts that only hits, commission misses and false alarms occur. For exhaustive searches, predicted response outcomes include these outcomes as well as omission misses and correct rejections. Model functions were derived which predict the probability of occurrence for these search and response outcome combinations.³ The resulting probability functions for the 11 predicted search strategy and response outcome combinations are given by equations 1, 2, 3, 4, 5, 6, 7, 8, 14, 15 and 18 listed in Table 16.

The predicted effects that variations in the model distributions have on discriminability among targets and alternatives are straightforward. In Figure 10, as the mean semantic distance value for the correct alternative distribution varies from low to high, the distribution shifts from right to left. The inverse relationship between semantic distance values and semantic similarity should be noted. The shift in the correct alternative distribution results in a decrease in discriminability due to the increased overlap of this distribution with the distribution representing incorrect alternatives. A similar effect can be produced as the mean semantic distance value for the incorrect alternative distribution varies from high to low. The model assumes that subjects adjust criteria to variations in menu item discriminability. However, the model does not assume that changes in discriminability resulting from variations in either distribution have an identical effect on criteria, or that changes in each distribution affect the criteria in a different manner. Resolution of this issue was determined empirically using procedures described later in this section.

Variations in the Number of Alternatives and Omission Probability are also predicted to affect criteria and response outcome probabilities. The overall trial probabilities of an incorrect menu alternative falling within a specific range of similarity values changes as a function of the Number of Alternatives. This results in an adjustment of criteria. Variations in Omission Probability have a similar effect on the overall outcome probabilities, resulting in adjustments to selection criteria.

Although selection criteria are predicted to adjust to variations in Alternatives, Omission Probability, and the two model distributions, there is no theoretical basis for predicting the specific nature of these adjustments. The effect that variations in these parameters have on model criteria was examined empirically using data from the menu task experiment described in Section III. Parameter estimates for the two distributions were used with values for the Number of Alternatives and Omission Probability to fit performance data observed for individual cells of the menu task experiment to model predictions. Best-fit solutions were derived by varying model criteria and other model parameters to minimize the differences between predicted and observed menu task performances. Once best-fit solutions were derived, relationships between model criteria and other model parameters were examined through a series of regression analyses.

Data-Fitting Methods

Predicted probabilities of menu task search strategies and response outcomes for variations in Alternatives, Omission Probability, and areas A through F under the two curves in Figure 10 are given by the probability functions listed in Table 16. Within the framework of these equations there are a number of ways in which predictions could have been derived. What follows is a description of two methods that were used in the present study. Both are derivations of fitting methods reported by Smith et al. (1974).

Method A

Method A used semantic similarity distances derived from the scaling procedures described in Section III to compute mean and standard deviation estimates for the two model distributions.

³A detailed summary of the procedures used to derive model equations appears in Pierce (1989).

**Table 16. Probability Functions Search Strategies by
Response Outcomes for the Two-Criterion Menu Model**

Number	Description	Function
1	P(ST Hit)	$[(1 \div a) F + (\sum_{i=1}^{a-1} (1 \div a) F(1-C)^i)] [1-Po]$
2	P(ST CM)	$[C \div a] [(a-1) + (\sum_{i=1}^{a-2} (a-i-1)(1-C)^i) +$ $(1-F)(1 + (\sum_{i=1}^{a-2} (i+1)(1-C)^i))] [1-Po]$
3	P(ST FA)	$[C] [1 + (\sum_{i=1}^{a-1} (1-C)^i)] [Po]$
4	P(Ex Hit)	$(E(A^{a-1})) [1-Po]$
5	P(Ex CM)	$[(a-1)DB(A^{a-2})] [1-Po]$
6	P(Ex FA)	$[aB(A^{a-1})] [Po]$
7	P(Ex OM)	$[D(A^{a-1})] [1-Po]$
8	P(Ex CR)	$[A^a] [Po]$
9	P(Red Search, CA Present and $C0 < Xc < C1$)	$[E \sum_{i=1}^{a-1} () A^{a-1-i} B^i] [1-Po]$
10	P(Red Hit $a = 2$ and CA present)	$\int_{C0}^{C1} \left[\int_{C0}^x e^{-(1/2)[y-\mu Xi]^2} dy \right] e^{-(1/2)[x-\mu Xc]^2} dx$
11	P(Red CM $a = 2$ and CA present)	$\int_{C0}^{C1} \left[\int_x^{C1} e^{-(1/2)[y-\mu Xi]^2} dy \right] e^{-(1/2)[x-\mu Xc]^2} dx$
12	P($C0 < m$ Xi values $< Xc < C1$ CA present)	$\int_{C0}^{C1} \left[\int_{C0}^x e^{-(1/2)[y-\mu Xi]^2} dy \right]^m e^{-(1/2)[x-\mu Xc]^2} dx$

Table 16. (Concluded)

Number	Description	Function
13	P(C0 < Xc < and C0 < m Xi values < C1 and at least 1 Xi < Xc CA present)	$\sum_{i=1}^m \binom{m}{i} \int_{C0}^{C1} \left[\int_x^{C1} e^{-(1/2)[y-\mu Xi]^2} dy \right]^i$ $\left[\int_{C0}^x e^{-(1/2)[z-\mu Xi]^2} dz \right]^{m-i} e^{-(1/2)[x-\mu Xc]^2} dx$
14	P(Red Hit)	$\sum_{i=1}^{a-1} \binom{a-1}{i} A^{a-i-1} \text{ (EQ 12; } m=i) [1-Po]$
15a	P(Red CM a = 2)	[EQ 13; m = 1] [1-Po]
15b	P(Red CM a > 2)	[(EQ 16) + (EQ 17)] [1-Po]
16	P(Red CM and C0 < Xc < C1 CA present)	$\sum_{i=1}^{a-1} \binom{a-1}{i} A^{a-i-1} \text{ (EQ 13; } m=i)$
17	P(Red CM and Xc < C0 CA present)	$D \left[\sum_{i=2}^{a-1} \binom{a-1}{i} A^{a-i-1} B^i \right]$
18	P(Red FA)	$P(\text{Red FA}) = \sum_{i=0}^{a-2} \binom{a}{i} B^{a-i} A^i [Po]$

Abbreviations.

In Descriptions.

ST - Self-terminating search
 Ex - Exhaustive search
 Red - Redundant search
 CM - Commission miss
 FA - False alarm
 OM - Omission miss
 CR - Correct rejection
 CA - Correct alternative
 IA - Incorrect alternative
 Xc - Semantic relatedness value of CA
 Xi - Semantic relatedness value of IA
 C0 - Criterion for immediate rejection
 C1 - Criterion for rapid selection
 Srch - Search
 P - Probability

In Functions.

A, B, C, D, E, F - Areas defined by
 distributions and criteria
 a - Number of alternatives
 Po - Probability of correct alternative absent
 μXc - Mean relatedness value for CA distribution
 μXi - Mean relatedness value for IA distribution
 EQ - Equation

Distribution estimates were computed separately for menu stimuli used in each cell of the menu task experimental design. The distribution representing the relationship between targets and correct alternatives was estimated for each cell by computing the mean and standard deviation of D_c values (scaled distance from target to correct alternative) associated with trials comprising that cell. Similarly, the distribution representing the relationship between targets and incorrect alternatives was estimated for each cell by computing the mean and standard deviation of the D_i values (scaled distances from correct to each incorrect alternative) associated with each incorrect alternative across trials in each cell.

Under Method A, distribution means and standard deviations, along with Alternatives and Omission Probability values for each cell, were fixed during computation of best-fit solutions. Values for the two criteria C_0 and C_1 were free to vary until differences between predicted and observed performance proportions were minimized.

Combined search strategy by response outcome proportions corresponding to the combinations predicted by the model equations listed in Table 16 were computed for individual subjects for individual cells using only the sequential Presentation Mode data from the menu task experiment. These proportions were operationally determined using the identical definitions of search strategy and response outcome described in Section IV. It should be remembered that the case where a subject examined all alternatives once and only once and chose the alternative in the last position in the menu was considered to be an indeterminate search. However, discarding these data, as was done with analyses of menu task search performances (see Section IV), would artificially increase cell proportions of redundant searches associated with the data set. Additionally, discarding these data would have affected the accuracy of computed response outcome proportions. Therefore, the decision was made to adjust observed self-terminating and exhaustive search data to account for the number of indeterminate search trials observed.⁴

Mean proportions for each of the 11 possible combinations of search strategy and response outcome were computed across subjects for individual cells and used as input to the left side of the model equations presented in Table 16 to derive estimates of the two criteria C_0 and C_1 . Best-fit solutions were derived by initially setting C_0 to the D_i distribution mean and C_1 to the D_c distribution mean, and then systematically varying criteria values until a best-fit criterion was met. The fitting procedure used an ordinary least-squares method where the sum of the square error values computed between predicted and obtained proportions across all prediction equations was minimized for each trial cell. Mean proportions from a total of 32 cells (representing all combinations of Alternatives, Omission, Hierarchy, and Nesting factor levels in the menu task experiment) were fitted for sequential Mode data.

⁴Because search strategy for indeterminate search trials could be considered either self-terminating or exhaustive, proportions of response outcomes for self-terminating and exhaustive searches were adjusted for individual subjects within each cell based on the number of indeterminate search strategy trials observed for each type of response outcome. To the extent that, by definition, indeterminate search strategies contained only responses where an alternative was chosen (i.e., hits, commission misses, or omission misses), only data associated with these types of responses were affected by this procedure. In the sequential Presentation Mode database, 1,318 of a total of 6,144 trials resulted in indeterminate search strategy.

For each subject, a ratio was computed for each cell of observed self-terminating search trials divided by the sum of both self-terminating and exhaustive search trials on which an alternative was chosen. The count of self-terminating search trials for each of the three response outcomes affected was then adjusted. This was done by adding, to the determinable number of self-terminating search trials observed for each of these response outcomes, the number of indeterminate search strategy trials observed for that response outcome type multiplied by this ratio. Similarly, for each subject, a ratio was computed of observed exhaustive search trials divided by the sum of both exhaustive plus self-terminating search trials on which an alternative was chosen. The count of exhaustive search trials for each of the three response outcomes was then adjusted by adding, to the determinable number of exhaustive search trials, the number of indeterminate search strategy trials observed for that response outcome multiplied by the ratio. Proportions for all search strategy and response outcomes were then computed by dividing the adjusted counts by the total number of trials performed for the cell.

Method B

Model-fitting procedures conducted under Method B were identical to those for Method A in all respects except the number of free parameters used to derive best-fit solutions. In addition to C0 and C1, the standard deviation parameters representing the Dc and Di distributions were also included as free parameters in Method B. Best-fit solutions were derived by initially setting values for these standard deviations to the values used under Method A, and then systematically varying these values along with C0 and C1 values until a best-fit criterion was met. The intent of this method was to determine the extent to which model-fitting improved as standard deviation parameters were allowed to vary. Improvement was assessed by comparing the root mean square of the differences between predicted and obtained performances across cells for each solution.

Procedures

A computer program was developed to compute separate fitting solutions for data obtained from each cell of the experimental design using Methods A and B. The program derived solutions for integrals associated with the equations in Figure 11 and Table 16 through numeric approximation.⁵ Best-fit values for free parameters were found by a hill-climb method. In the hill-climb, free parameters were sequentially incremented by a delta. After each increment, the fit was evaluated by summing the squared deviations between predicted and observed proportions. If the fit improved, the parameter was set to the new value. If the fit did not improve, the original value of the parameter was decremented by the delta and the fit was recomputed. If this resulted in a fit improvement, the parameter was set to the decremented value. If the fit did not improve, the parameter was set to its original value. This was repeated for each free parameter, in turn, until a pass was made that resulted in no fit improvement. At this point, the delta was decreased and the process was repeated until the delta reached a preset minimum.

The program computed fitting solutions using Methods A and B for several subsets of the database in order to address certain problems associated with the design of the menu task experiment.⁶ It may be recalled that analysis of Omission Probability manipulation effects on

⁵Because of the inverse relationship between semantic distance and semantic similarity, the computation of numeric approximations for integrals associated with equations in Figure 11 and Table 16 required that all distribution means that were estimated for each of the 32 cells be transformed to negative values (i.e., multiplied by -1). Thus, an estimated distribution mean of 1.5 was transformed to -1.5. This had the effect of reversing the original relationship between distance values and semantic relationships. For example, while a change in the Dc mean from 1.5 to 1 prior to the transformation reflected a decrease in semantic distance, it also reflected an increase in semantic similarity. After transformation of these values, the change from -1.5 to -1 resulted in both an increase in the transformed value as well as an increase in semantic similarity. This transformation procedure resulted in derivation of criteria estimates whose values also related directly to levels of semantic similarity. Upon completion of computations to derive best-fit solutions, the same transformation was conducted a second time on values for distribution means and model criteria. Thus, all values for distribution means and model criteria reported in the text reflect values that are inversely related to semantic similarity; i.e., increases in values for distribution means or criteria reflect decreases in semantic similarity.

⁶In addition to those fitting solutions reported in the main body of this text, Method A and B fitting solutions also were computed across all trials for each cell using estimates of distribution parameters for Correct Alternative Present trials only. In these solutions, data in all but 11 cells were fitted to equations 1, 2, 3, 4, 5, 6, 7, 8, 14, 15, and 18 in Table 16. In cells where Omission Probability equaled zero, and in cells where Number of Alternatives equaled eight and the Hierarchy factor level was close, all trials had menus where the Correct Alternative was Present. For these 11 cells, data were fitted only to the 7 model equations that predicted probabilities associated with trials where the Correct Alternative was Present. Thus, probability functions for false alarm and correct rejection outcomes were omitted from these solutions and only equations 1, 2, 4, 5, 7, 14, and 15 in Table 16 were used.

In addition, separate Method A and B solutions were computed for Correct Alternative Absent data using distribution parameter estimates for Correct Alternative Absent trials using equations 3, 6, 8, and 18 in Table 16. Because of the limited number of observations per cell for Correct Alternative Absent trials (ranging from zero to twelve), estimates of the Di distribution parameters used in computing these solutions were questionable.

Resulting fits for the all trial solutions were highly similar to those obtained for the Correct Alternative Present trial solutions reported in the text. However, solutions for the Correct Alternative Absent trial data were relatively poor and not readily interpretable. Due to the problems associated with both of these sets of solutions, discussion in the text was limited to only Method A and B solutions where proportions for only Correct Alternative Present trials were fitted.

Dc and Di values in the menu task experiment showed significant differences in mean Di values between Correct Alternative Present and Correct Alternative Absent trials. These results suggest that the Di distributions for Correct Alternative Present and Absent trials were not the same. Though separate estimates of Di distribution parameters for the Correct Alternative Absent trials were available, the limited number of trials per cell (ranging from 0 to 12) that were available to estimate these parameters made the reliability of these estimates questionable. However, a relatively clean data set for which large numbers of trials per cell were available was obtained using data from Correct Alternative Present trials. Therefore, discussion will focus only on Method A and B solutions derived for all cells using data from trials only where the Correct Alternative was Present. For these solutions, probability functions for false alarms and correct rejection outcomes were omitted and only equations 1, 2, 4, 5, 7, 14, and 15 in Table 16 were used.

Another problem that needed to be addressed concerned the fact that the model predicted zero choice responses (none of the alternatives are chosen as correct) only under conditions where an exhaustive search was made. Whereas over 90% of zero choice responses were made after an exhaustive search of the alternatives, 106 out of a total of 1,106 zero choice responses were made after a redundant search. These responses were not equally distributed across cells. Thus, in order to derive best-fit solutions with comparable measures of fit across all cells, redundant search trials where zero choice responses were made were deleted from the database. The remaining data sets were then normalized such that the proportions for the search-strategy-by-response-outcome categories summed to one.

In summary, proportions of search strategies and response outcomes for Correct Alternative Present trials observed for each of the 32 design cells were used to derive criteria estimates using Method A and B fitting techniques. Thus, Correct Alternative Present data were fitted to equations 1, 2, 4, 5, 7, 14, and 15. These equations correspond only to search-strategy-by-response-outcome combinations predicted when the Correct Alternative was Present in the menu. Distribution parameters were derived from the scaled estimates of menu stimuli semantic distances for Correct Alternative Present trials in each cell. Finally, all observed proportions for each cell were normalized to sum to one.

Findings

Model Solutions

Root mean squared error (RMSE) values and correlations between predicted and obtained search-strategy-by-response-outcome proportions were used as indicants of the fit between empirical data and model predictions. RMSE values represent the root mean of the squared differences between predicted and observed proportions. Tables listing model parameter and RMSE values that were derived from Method A and B solutions for individual cells are given in the Appendix. In the tables in the Appendix, criteria values represent semantic distance measures and, therefore, have the same inverse relationship to semantic similarity as do Dc and Di values. As values for either criterion increase, the degree of semantic similarity required to exceed the criterion decreases. Also included in the Appendix are RMSE and correlation values computed between observed and predicted proportions for individual equations across cells.

RMSE values computed across all seven equations and all cells were .079 and .018 for Method A and B solutions, respectively. Correlations between predicted and observed proportions across all seven equations and all cells were .887 for the Method A solutions and .994 for the Method B solutions. Though the data used to compute these correlations were not independent (within each cell, the sum of the proportions for all equations equaled one), the correlations are reported here to serve as a measure of congruency between predicted and observed proportions.

Effects of Menu Design Factors on Model Criteria

The effects that variations in the Number of Alternatives, Omission Probability, and the means and standard deviations of the two model distributions had on model criteria were examined through a series of regression analyses. Data used for these regressions were derived from the best-fit model solutions of menu task performance data. A total of four regressions were computed using data derived for individual cells as the basic unit of analysis. In two of the regressions, the value of the rapid rejection criterion, C0, that was derived for individual cells was used as the criterion measure. In the other two regressions, the criterion measure used was the value of the rapid selection criterion, C1, derived for individual cells. For each of these two criterion measures, one regression was computed using parameter estimates from the Method A solution and another regression was computed using parameter estimates from the Method B solution. In all four regressions, predictor variables were Number of Alternatives, Omission Probability, the Dc mean, the Dc standard deviation, the Di mean, and the Di standard deviation computed for each cell.

Results of regressions for the rapid rejection criterion, C0, are summarized in Table 17 for both Method A and B solutions. Significant F-values were obtained for both regressions. R-squares for these regressions indicated that over 95% of the C0 variance was accounted for by model predictors.

Comparison of R-square values between the two C0 regressions indicated a high degree of consistency. In addition, signs of the significant partial coefficient estimates were consistent across the two regressions. These findings notwithstanding, the significance of the partial correlations and coefficients for individual predictors differed somewhat across the two regressions. In regressions computed for the Method A solution, all predictors with the exception of the Di standard deviation significantly contributed to the regression model, and all but the Number of Alternatives predictor significantly contributed to the regression model for the Method B solution.

Results of regressions for the rapid selection criterion, C1, are summarized in Table 18 for the Method A and B solutions. Significant F-values were obtained for both regression models. R-squares were .49 for the Method A data and .74 for the Method B data.

There were substantial differences obtained between the C1 regression results for Method A and B solution data. Regression models computed for Method B data accounted for over 24% more variance than did regressions for Method A data. Across both regressions, the partial correlations and coefficients for the Number of Alternatives predictor were significant. However, whereas regression results for the Method A solution also indicated that the partial correlations and coefficients for the Di standard deviation were significant, these statistics were not significant for this predictor in the regression results for Method B solutions. Finally, regressions computed for Method B solution data indicated that the partial correlations and coefficients for the Dc mean were significant. In contrast, these statistics were not significant in regressions of Method A solution data.

Synopsis

In the absence of a well-defined criterion measure or competing model whose predictions could be used for comparison to the results obtained for the two-criterion menu model, it is difficult to evaluate goodness of fit between predicted and observed task performances. Even so, RMSE and correlation values between model predictions and observed proportion data did provide a means of comparing differences in fits derived for the two fitting methods used.

Differences between regression results for the two fitting methods were obtained for both criterion measures. Whereas in Method A only model criteria were free to derive best-fit solutions,

Table 17. Regression Results for Rapid Rejection Criterion (C0) Estimates from Method A and B Solutions

		Method A Solution				Method B Solution			
ANOVA results		6-Predictor Model				6-Predictor Model			
Source		DF	SS	MS	F	DF	SS	MS	F
Model		6	6380	.1063	107.149***	6	1.1597	.1933	93.982***
Error		25	.0248	.0010		25	.0514	.0021	
Total		31	.6628			31	1.2111		
		R-square = .9626				R-square = .9575			
Predictor variables		pr	Coef	Std Err	t	pr	Coef	Std Err	t
Intercept			.166	.064	2.598*		.151	.055	2.753*
Alternatives		.738	-.022	.003	-8.390***	.009	-.002	.004	-.472
Omission Prob		.262	-.120	.040	-2.978**	.202	-.174	.069	-2.514*
Dc Distribution a		.404	.305	.074	4.116***	.267	.290	.096	3.020**
Dc Distribution b		.452	1.263	.278	4.545***	.186	.359	.150	2.392*
Di Distribution a		.728	.256	.031	8.173***	.675	.616	.085	7.213***
Di Distribution b		.004	.042	.127	.327	.573	-1.052	.182	-5.796***

Note: All 32 cells are included in these regressions. Data were obtained from best-fit solutions for 7 Correct Alternative. Present equations only.

*p < .05

**p < .01

***p < .001

Table 18. Regression Results for Rapid Selection Criterion (C1) Estimates from Method A and B Solutions

ANOVA results		Method A Solution					Method B Solution				
Source		6-Predictor Model					6-Predictor Model				
		DF	SS	MS	F		DF	SS	MS	F	
Model		6	.9473	.1579	3.973**		6	.2487	.0414	11.558***	
Error		25	.9935	.0397			25	.0897	.0036		
Total		31	1.9408				31	.3384			
		R-square = .4881					R-square = .7350				
Predictor variables		pr	Coef	Std Err	t		pr	Coef	Std Err	t	
Intercept			1.280	.404	3.169**			.157	.072	2.165*	
Alternatives		.355	.062	.017	3.709**		.355	.021	.006	3.711***	
Omission Prob		.001	.040	.256	.157		.022	.069	.091	.755	
Dc Distribution μ		.098	-.772	.470	-1.644		.389	.506	.127	3.993***	
Dc Distribution σ		.001	-.324	1.758	-.184		.059	-.248	.198	-1.253	
Di Distribution μ		.074	.280	.198	1.413		.004	.037	.113	.329	
Di Distribution σ		.221	-2.137	.803	-2.662*		.001	.032	.240	.132	

Note. All 32 cells are included in these regressions. Data were obtained from best-fit solutions for 7 Correct Alternative Present equations only.

*p < .05.

**p < .01.

***p < .001.

in Method B both the model criteria and the standard deviation parameters were free to derive fitting solutions. Though it is tempting to attribute the better fits obtained for Method B solutions solely to the optimization of standard deviation estimates for the two model distributions, such a conclusion would be difficult to support. Similar improvement in fits could have been demonstrated by fixing the standard deviation parameters and freeing up parameters representing the means of the two model distributions. Thus, improvements in the Method B fits could be attributed not only to better estimation of the standard deviation parameters, but also to improvements that could have been derived from better estimation of distribution means. Finally, it could be argued that the observed improvements in fit may have been due simply to a better accounting of error between predicted and obtained proportions that cannot be attributable to improvement in the accuracy of estimating any of the model parameters. Although these results suggest the possibility that there may exist better measures that could be used to represent model distribution parameters, the difficulty in identifying the real source of these fit improvements leaves resolution of this issue to further research.

The issue raised above has further implications concerning interpretation of the results of regressions computed for the two model criteria. Because the standard deviation estimates derived from Method B solutions cannot be interpreted conclusively as reflecting values that represent the actual standard deviations of these distributions, the relationship between regression model predictions and the two criteria is clearly better reflected by regressions performed on the data derived from Method A.

Focusing on the partial coefficients tested in this latter set of regressions, the rapid rejection criterion, C0, was shown to be significantly affected by variations in all predictors except the one representing the Di standard deviation. C0 increased (reflecting a decrease in the degree of semantic similarity required to exceed it) as the Number of Alternatives and Omission Probability decreased. C0 also increased as the Dc standard deviation and both distribution means increased. The rapid rejection criterion, C1, was significantly affected only by variations in the Number of Alternatives and the Di standard deviation. As Number of Alternatives increased, C1 increased. As the Di standard deviation increased, C1 decreased. It appears from these results that while all factors contributed to variations in menu criteria, their individual effects were more or less pronounced depending on the criterion in question.

These differences are further exemplified through comparison of the effects that distribution parameters had on the two criteria. Results from the C0 regression using Method A data suggest that variations in the Di and Dc distributions had similar effects on the C0 criterion. Thus, it would appear that overall discriminability among alternatives had a significant effect on the value of C0. Results from the C1 regression indicate that whereas variations in the Di standard deviation significantly affected values of the C1 criterion, Dc distribution parameters did not. In contrast to the results obtained for C0 regression, these results suggest that C1 was not affected as much by the overall discriminability among alternatives (represented by variations in either of the model distributions) as by variations in the Di distribution alone.

In summary, the fitting procedures and regression analyses provided a means of assessing the extent to which data could be fitted to model equations, as well as a means of assessing the relationship between menu design factors and model criteria. Comparison of the results of the two methods used to compute fitting solutions suggested further research as to how model distributions might best be estimated. Finally, examination of regression results obtained for the Method A solution indicated that variations in model criteria are significantly related to variations in Number of Alternatives, Omission Probability, and the estimated means and standard deviations of model distributions.

VI. DISCUSSION

The previous section described the two-criterion menu model and procedures by which empirical data were fitted to the model to derive estimates of model criteria for cells of the

experimental design. Regressions on criteria values derived from these procedures showed that significant relationships exist between model criteria and menu design factors. The present section uses these results to explicate the results of the menu task performance analyses described in Section IV, and then proposes an agenda for addressing relevant, unresolved issues through future research.

Explication of Menu Task Performance Results

As described in the previous section, model criteria values were derived only for menu task data that were obtained for sequential Presentation condition trials in which the Correct Alternative was Present. Discussion will be limited, therefore, to performance results associated with these data. These results pertain to proportions of hits, self-terminating searches, exhaustive searches, and redundant searches for trials performed under the sequential Mode condition where the Correct Alternative was Present. In addition, because model probability functions were based on semantic distance distributions estimated for all alternatives in each menu, only performance results for models containing the D_i average predictor are considered. The effects that each of the four menu design factors (Number of Alternatives, Omission Probability, semantic distance between target and correct alternative, and average semantic distance between correct and incorrect alternatives) had on model criteria and model predictions are each considered separately. The effect that each factor had on menu task performance is then explained in terms of the processes suggested by the model.

The relationships between model criteria and areas under the model distribution curves were presented in Figure 10. To aid the discussion of these relationships, Figure 12 provides a summary of results from menu task performance and model criteria regressions. The partial coefficients for the four menu design factors that were derived from regressions of sequential mode hit and search strategy proportions (originally presented in Tables 10 and 15) are presented in Figure 12a. Figure 12b lists the partial coefficients computed from the regressions on model criteria values that were derived from the Method A fitting procedure (originally presented in Tables 17 and 18).

Effects of Variations in Number of Alternatives

Model parameter relationships. The curves shown in Figure 10 represent the distributions of relatedness values among targets and correct alternatives, and among targets and incorrect alternatives. As the number of incorrect alternatives in the menu increases, the greater the likelihood that at least one of the incorrect alternatives in the menu will fall within any one of the three areas defined under the upper curve in this figure. Thus, as the Number of Alternatives increases, the greater the probability that at least one incorrect alternative will exceed both the value of the correct alternative and the values of model criteria.

Additionally, as the Number of Alternatives increases, subjects require increasing relatedness between target and alternatives to satisfy requirements of the C_0 criterion. This finding is reflected in Figure 12b by decreases in the value of C_0 as Number of Alternatives increases. Because the values associated with both criteria were set to the same scale that was used for values of D_c and D_i , decreases in criteria values reflect increases in semantic relatedness requirements. Conversely, the relatedness requirements to exceed the rapid selection criterion decrease as the Number of Alternatives factor increases (reflected in Figure 12b by increasing C_1 values).

Effects on Response Accuracy. As the Number of Alternatives increases, the model predicts fewer responses that result in a hit, due to (a) an increase in the likelihood that at least one

Response Accuracy:		Hits - Sequential mode only	
PREDICTOR VARIABLES		Coef	Std Err
Alternatives		-.062	.012
Omission Prob		-.908	.195
Dc Distance		-1.042	.112
Di AVE Distance		.403	.084

Search Strategy:	Self-terminating searches				Exhaustive searches				Redundant searches			
	Coef		Std Err		Coef		Std Err		Coef		Std Err	
PREDICTOR VARIABLES	Coef	Std Err	t		Coef	Std Err	t		Coef	Std Err	t	
Alternatives	.063	.013	4.729***		-.103	.012	-8.763***		.075	.010	7.139***	
Omission Prob	.554	.205	2.710***		-.052	.181	-.289		-.745	.161	-4.639***	
Dc Distance	-1.241	.124	-10.019***		.741	.109	6.772***		.637	.097	6.545***	
Di AVE Distance	.052	.096	.539		.083	.084	.986		-.264	.075	-3.516***	

a. Performance regression results

C0 Criterion				C1 Criterion			
PREDICTOR VARIABLES		Coef	Std Err	PREDICTOR VARIABLES		Coef	Std Err
Alternatives		-.022	.003	Alternatives		.062	.017
Omission Prob		-.120	.040	Omission Prob		.040	.256
Dc Distribution μ		.305	.074	Dc Distribution μ		-.772	.470
Dc Distribution σ		1.263	.278	Dc Distribution σ		-.324	1.758
Di Distribution μ		.256	.031	Di Distribution μ		.280	.198
Di Distribution σ		.042	.127	Di Distribution σ		-2.137	.803

b. Method A modeling regression results

*p < .05.
 **p < .01.
 ***p < .001.

Figure 12. Comparison Between Performance Data Results and Model Predictions.

incorrect alternative will exceed the relatedness value of the correct alternative in the menu (thus resulting in a commission miss) and (b) an increase in the likelihood of an omission miss as a result of increases in semantic relatedness requirements associated with the C0 criterion. With both commission misses and omission misses predicted to increase with increasing Numbers of Alternatives, proportions of a hit must decrease. The regression results for proportions of hits are consistent with model predictions. As indicated in the summary of response accuracy regression results listed in Figure 12a, proportions of hits decreased as Number of Alternatives increased.

Effects on Search Strategies. The effects of increases in Number of Alternatives on search strategies employed by subjects during menu task performance are also predicted by the model. The regression results for the C1 criterion indicate that relatedness requirements to meet this threshold decreased as Number of Alternatives increased. Thus, the model predicts that increases in Numbers of Alternatives results in lower semantic similarity requirements to produce a self-terminating search. In addition, with more incorrect alternatives in the menu, the greater the probability that at least one of them will meet the rapid selection threshold. The effect of increased Number of Alternatives obtained from analysis of self-terminating search performance data are consistent with these predictions. As indicated in the summary of search strategy regression results listed in Figure 12a, as Number of Alternatives increased, self-terminating searches significantly increased.

The results obtained from analysis of exhaustive and redundant search performance data for variations in the Number of Alternatives are also consistent with model predictions. The regression of model parameter data on the C0 criterion indicates that the semantic relatedness requirements to meet this threshold increased with increasing Numbers of Alternatives. Furthermore, as Number of Alternatives increased, the relatedness requirements to meet the C1 rapid selection criterion decreased. These results indicate a decrease in the likelihood that any one particular alternative will have a relatedness value that falls between criteria. To some extent, however, this tendency is offset by the predicted increase in the likelihood that with more Alternatives, the probability of having more than one alternative fall within any given range of relatedness values increases. Thus, a tradeoff between increased Numbers of Alternatives competing for selection versus a decrease in the range of values that define the area under the distributions between criteria is suggested by the model.

Results of the analysis of exhaustive and redundant search data indicate that increases in the Number of Alternatives resulted in an increase in the overall probability of having more than one alternative fall between criteria. Figure 12a indicates that as Number of Alternatives increased, exhaustive search proportions decreased and redundant search proportions increased. Thus, the decreases in Areas B and E under the probability curves in Figure 10 that result from the adjustments in criteria are exceeded by overall probability gains that result from increases in the Number of Alternatives competing for selection.

Effects of Variations in Omission Probability

Model Parameter Relationships. Because all the results listed in Figures 12a and b were computed for trials where the Correct Alternative was Present, Omission Probability could affect menu task performance only as mediated through the effect it had on the two model criteria. As indicated in Figure 12b, Omission Probability significantly affected variations in the rapid rejection criterion, C0, but did not have a significant effect on the C1 rapid selection criterion. These results indicate that as Omission Probability increased, a stronger semantic relationship between target and alternatives was needed to satisfy requirements of the C0 criterion.

Effects on Response Accuracy. A significant effect due to variations in Omission Probability was obtained for the proportions of response outcomes resulting in hits. As Omission Probability increased, proportions of hits decreased. This effect is explained in terms of the model by the effect Omission Probability was shown to have on the C0 rapid rejection criterion. As Omission Probability increased, the semantic relatedness requirements to meet this threshold increased. As a result, the probability of the correct alternative having a semantic distance value that could not meet the C0 criterion increased with increasing Omission Probability. This resulted in an increase in the likelihood of immediately rejecting the correct alternative as a response candidate. Thus, the probability of a hit decreased at the expense of an increase in the probability of omission misses.

Effects on Search Strategies. Because C0 was the only criterion that was influenced significantly by variations in Omission Probability, the model predicts that the only types of searches that should be affected by variations in Omission Probability are redundant and exhaustive. More specifically, due to the fact that semantic relatedness requirements to meet the C0 criterion significantly increased as Omission Probability increased, the model predicts that the probability of having more than one alternative fall between criteria decreases. Thus, exhaustive searches are predicted to increase at the expense of decreases in redundant searches. Further, because the partial coefficient for the Omission Probability predictor was not significant in the regression on C1 estimates, the probability that at least one alternative will satisfy the rapid selection criterion should remain relatively stable across varying levels of Omission Probability. Therefore, self-terminating search proportions are predicted to remain relatively constant across varying levels of Omission Probability.

The results of search strategy regressions summarized in Figure 12a indicate that redundant searches decreased as Omission Probability increased, as predicted by the model. The results obtained for self-terminating and exhaustive searches, however, were contrary to model predictions. Omission Probability had no significant effect on proportions of exhaustive searches. Further, self-terminating searches increased as Omission Probability increased.

It is evident that the relationships between Omission Probability and search strategies indicated through the analysis of performance data are inconsistent with processes suggested by the model. Analyses of performance data suggest that as Omission Probability increased, the relatedness requirements associated with the C1 criterion decreased. This decrease in C1 requirements, combined with the increase in C0 requirements that was indicated through the regression on model-fitting data, suggests a process that is consistent with all obtained performance results. Conceptually, this process proposes that subjects required greater relatedness of alternatives for selection consideration; however, once the relatedness exceeded this threshold, subjects required lower relatedness of alternatives for immediate selection. Exactly why the results from the data-fitting regressions on the C1 criterion were contrary to processes suggested by the analysis of search strategy performances cannot be ascertained at this time. It is suspected that differences in the units of analysis and resulting differences in the power of the two sets of analyses may have contributed to this inconsistency of results.

Effects of Variations in Dc

Model Parameter Relationships. Method A procedures for deriving criteria values were based on estimates of Dc and Di distribution parameters computed across sets of trials as defined by the cells of the experimental design. Thus, the partial coefficients for the Dc and Di parameters summarized in Figure 12b illustrate the relationships between variations in model criteria and distribution parameters across sets of trials where Alternatives and Omission Probability within each set were held constant. The performance regression results listed in Figure 12a were based on performances for individual trials across all cells of the design. Whereas Number of

Alternatives and Omission Probability were held constant for trials within each cell, Dc and Di values varied from trial to trial. Therefore, the use of model criteria regression results in explaining the effects that variations in Dc and Di had on menu task performances is complicated by the use of different units of analysis for these two sets of regressions. In addition, the relationship between variations in the standard deviation of distance values across sets of trials and variations in the actual distance values from trial to trial is unknown. For these reasons, it is difficult to map the effects on criteria that were due to differences in the standard deviations of the two model distributions onto the effects that distance variations had on menu task performances. These problems notwithstanding, the explanation of the Dc and Di effects on task performance will be discussed by initially considering model predictions for the case where Dc and Di vary across relatively stable criteria, and secondly, by considering model predictions for variations in distance means across sets of trials.

Considering the criteria regression results of the analyses for the Dc parameters in Figure 12b, significant partial coefficients were obtained for variations in Dc distribution means across sets of trials in the regression on the C0 criterion. However, partial coefficients for Dc parameters were not significant in regressions on the C1 criterion. These results indicate that as Dc means increased across sets of trials, the only significant change observed for values of model criteria was a decrease in semantic relatedness requirements associated with the C0 criterion.

Effects on Response Accuracy. On any given trial, as the Dc value increases, the lower the probability will be that the correct alternative will have a semantic similarity to the target that exceeds the similarity between incorrect alternatives and the target. Thus, for stable criteria, increases in Dc are predicted to decrease proportions of response outcomes that result in a hit. Across sets of trials, as the Dc means increased, relatedness requirements associated with the C0 criterion significantly decreased. While allowing correct alternatives with a greater range of Dc values to be considered for selection, the change in C0 requirements does nothing to offset the decreased level of selection competition produced by the correct alternative. Thus, decreases in hit probability are also predicted for increases in Dc means across sets of trials. As indicated in Figure 12a, regression results obtained for hit performances are consistent with these predictions. Across all trials, as Dc values increased, the proportion of hits decreased.

Effects on Search Strategy. For any given trial, as Dc increases, the probability that the relatedness of the correct alternative will exceed relatedness requirements set by the rapid selection criterion decreases. Thus, for stable criteria, increases in Dc are predicted to decrease proportions of self-terminating searches. As indicated in Figure 12b, increases in Dc means were shown to decrease relatedness requirements set by the rapid rejection criterion (as indicated by increases in values of C0). This has the effect of increasing the number of alternatives that fall between criteria. The combination of these events results in predicted increases in proportions of both exhaustive and redundant searches at the expense of decreases in self-terminating searches. The search strategy regression results indicated in Figure 12a were consistent with these predictions. Across all trials, as Dc averages increased, self-terminating searches decreased and proportions of exhaustive and redundant searches increased.

Effects of Variations in Di

Model Parameter Relationships. As shown in Figure 12b, the effect of variations in Di means across sets of trials was significant for the regression on the C0 criterion, but was not significant for the C1 regression. As the Di means increased, the semantic relatedness requirements associated with the C0 criterion decreased. Conversely, the effect of variations in the Di standard deviation across sets of trials was not significant for the C0 regression, but its effect was significant for the regression on the C1 criterion. As noted during discussion of the effects of variations in Dc, the comparison between the effect that variations in Di had on menu task

performance and the effect that D_i had on model criteria is complicated by the fact that the two sets of regressions were based on different units of analysis. The explanation of performance results for variations in D_i , therefore, will follow the same theme that was used to discuss the effects of variations in D_c . Predicted effects of variations in D_i initially will be considered for the case where D_i varies across relatively stable criteria. Secondly, model predictions will be derived for variations in D_i means across sets of trials.

Effects on Response Accuracy. For any given trial, as D_i increases, the probability that the relatedness of any one of the incorrect alternatives in the menu will exceed the relatedness of the correct alternative decreases. Thus, for stable criteria, increases in D_i are predicted to increase the proportion of response outcomes resulting in hits. Across sets of trials, as the D_i means increase, the relatedness requirements of the C0 criterion decrease. Although this results in allowing incorrect alternatives with a greater range of D_i values to be considered for selection, the change in C0 requirements does nothing to offset the decreased level of selection competition produced by incorrect alternatives. Thus, increases in hit proportions are also predicted for increases in D_i means across sets of trials. The results obtained for the regression on proportion of hit responses shown in Figure 12a are consistent with model predictions. Across all trials, the proportions of hits increased as a function of increases in D_i averages.

Search Strategy Performances. As shown in Figure 12a, the only significant partial coefficient obtained for the D_i predictor was in the regression on redundant search proportions. As D_i average distances increased, the proportion of redundant searches decreased. Examination of Figure 10 indicates that incorrect alternatives having high D_i values are predicted to have a low probability of satisfying relatedness requirements associated with the C0 criterion. Thus, alternatives having high D_i values are predicted to be immediately rejected. The greater the probability of immediately rejecting incorrect menu alternatives, the lower is the probability that more than one alternative in the menu will be considered for selection. The effect of this process is offset to some extent for increases in D_i means across sets of trials. As indicated in Figure 12b, the relatedness requirements of the C0 criterion decrease with increasing D_i means. Menu task performance results, however, suggest that the decrease in C0 relatedness requirements is not sufficient to counter the increased number of incorrect alternatives that are immediately rejected when D_i values are high. This process results in the obtained effect that as D_i averages increase, redundant search proportions significantly decrease.

Summary

Analysis of menu task performance data showed that menu design factors significantly influence both search strategy and response accuracy. The two-criterion menu model provided a theoretical framework of menu search and decision processes that was employed to explain variations in menu task performances. Through the fitting of menu task data to prediction equations derived for the menu model, the relationship between model criteria and menu design parameters was assessed. The discussions presented in this section demonstrated how the processes suggested by the model accommodate most of the findings derived from analysis of menu task performance data. An exception to the consistency between model predictions and menu task performance was noted for the relationship between variations in Omission Probability and proportions of search strategies employed by subjects. Though the model predicted the results obtained for Omission Probability effects on redundant searches, the Omission Probability effects that were indicated for self-terminating and exhaustive searches were contrary to model predictions. This fact notwithstanding, performance results obtained for all other menu design factors studied in this investigation were consistent with processing suggested by the model.

Although the results of the present investigation contribute to the understanding of menu search and decision processes, several issues remain unresolved. What follows is a summary

of the contribution of the present work to the problem of modeling requirements for optimal menu design. Issues to be resolved through further study are discussed, and a research agenda is proposed for the development of an operational model that will provide menu designers the information necessary to optimize specific menu designs.

Research Agenda

Successful implementation of menu-driven systems across a variety of conditions requires a model in which all factors affecting search strategy and response accuracy are included. The results of the present investigation indicate the necessity of including Number of Alternatives, Omission Probability, and semantic relatedness between target and menu alternatives in such a model. Processes proposed under the two-criterion menu model meet this requirement, and were shown to accommodate many of the findings obtained from analysis of menu task performance data. Although the results of the model-fitting procedures described in Section V are relevant to the understanding of menu search and decision processes, the procedures fall short of a totally rigorous test of the model. The testing issue is left to future research. In addition, refinement of the model for use as an operational tool to resolve menu design issues for specific applications would also require further work.

Such effort might begin by further studying the problem of deriving a priori estimates of model criteria. Estimation of reliable menu criteria values through the linear combination of menu design parameters certainly appears feasible following from the results of the model-fitting analyses. However, the stability of model criteria relationships to menu design parameters across a variety of menu task situations needs to be determined. Further work is also required concerning the optimal manner by which model distribution parameters might best be estimated. Finally, tests designed to assess the predictive value of the model for several different types of menu stimuli need to be conducted to validate the accuracy of the model.

Once the model and its assumptions have been validated (and/or modified), the logical conclusion of this work would be to refine the model into an operational tool to be used for optimization of specific computer menu designs. This goal could be achieved through extending the model's predictive capabilities from its current focus, which concerns partial search menu task performances, to the general activity of navigation through menu hierarchies. This would require consideration of breadth/depth performance tradeoffs through computation of model parameter variations, and assessment of resulting variations in model performance predictions.

Several implications were derived from the present effort with respect to the significance of menu design factors on task performance, and with respect to the modeling of search and decision processes involved during the performance of the partial search menu task. Results demonstrated the inadequacy of menu optimization models which are based solely on subsets of the design factors studied here. Both the literature and the results of the present experiment suggest a complex processing phenomenon involved in menu search and decision making which involves all factors studied. Though falling short of a definitive test, the procedures used in the present investigation demonstrated the two-criterion menu model to be capable of fitting empirically obtained performance data in a manner that contributes to the understanding of the relationships between menu design factors and task performance.

VII. CONCLUSIONS

Results of the present investigation showed the significant effects that Number of Alternatives, Omission Probability and semantic relationships among menu stimuli had on menu task performance. It was also shown how variations in Number of Alternatives across levels of a

menu hierarchy were confounded by variations in D_c and D_i distances. These results indicated the complexity of the relationship that exists between variations in these factors and search strategy and response outcome performances. These findings suggested that in order to best understand this complex relationship, a framework that accounts for the effects of all these factors is required. Such a framework was developed through the proposed two-criterion menu model. The model was based on the relationship that exists between the target and menu alternatives for menus of any size and on the probability that the Correct Alternative is Present in the menu. Although it was shown that processes proposed under this model could be used to account for many of the findings obtained from analysis of performance data, these procedures fell short of a definitive test of the model's validity. Still, there are several implications that can be drawn from this work and used to form general recommendations for optimizing the structure of computer menus.

To optimize the menu structure, the designer must consider the effects that each menu in the hierarchy has on response speed and accuracy. As was indicated in the Section II review of the literature, the sum of these effects across all levels of the hierarchy has traditionally been used as the basic unit of measurement for determining optimum menu design. Optimal structure has been discussed typically in terms of designs in which response time is minimized or response accuracy is maximized, or in terms of tradeoffs between response speed and accuracy across all levels of the menu structure.

With regard to menu optimization studies, a primary issue has been the Number of Alternatives that should be incorporated in the design of the menu structure. The present investigation studied this issue in terms of factor effects on performance for individual menus that were derived from various levels of a menu structure. The results showed that Number of Alternatives was not only a significant factor in terms of its partial correlation to response accuracy, but also significantly confounded by variations in the relationship between correct and incorrect alternatives (i.e., D_i values). In addition, the strength of the relationship between Number of Alternatives and D_i variations was also shown to vary across levels in the menu hierarchy.

The implication these findings have with respect to optimum Number of Alternatives is that it is highly probable that this number will vary across menu levels for any given hierarchy. Because increases in Number of Alternatives differentially affect D_i -NN and D_i -AVE values for varying hierarchy levels, the tradeoffs among the influences each of these factors has on task performance do not appear to be consistent across levels in the hierarchy. In terms of the two-criterion menu model, this process is further complicated by variations both in D_c distributions and in criteria employed by subjects across levels of the menu structure. Thus, while the relationships among menu alternatives are dependent both on the hierarchical level from which alternatives are derived and on the Number of Alternatives contained in the menu, the overall performance effect of variations in Number of Alternatives is also dependent on selection thresholds. It is concluded that the optimum Number of Alternatives is dependent on the relationships among all model parameters for specific levels in the hierarchy from which menus are constructed.

In summary, though the development of an operational model that could be applied to the optimization of specific menu applications requires additional effort, the two-criterion model developed during this investigation provides both a quantitative and theoretical framework to guide the continuation of this work. Additionally, results of the experiment reported in this report indicate the complexity of the relationships among menu design factors and of their effects on task performance. Although work reported in the human factors literature has searched for an optimum Number of Alternatives that could be used to guide menu designers, results of this work appear to be confounded by the failure to consider the effects that Number of Alternatives has on semantic relationships among menu stimuli.

Results of the present investigation suggest that the optimum menu structure may be one in which the Number of Alternatives varies as the user travels through the menu hierarchy. The optimum Number of Alternatives for a given level of the hierarchy is not only expected to vary from one level to the next but, due to varying relationships among menu design factors, also expected to vary from one menu application to another. This optimum is dependent on the effects that all model parameters have on task performance at each level of the menu structure. Thus, in deciding upon the Number of Alternatives to use for a specific level of the hierarchy, the menu designer must consider the effect this number will have on menu breadth/depth, model criteria, and the relationships among targets and menu alternatives.

REFERENCES

- Barsalou, L.W. (1983). Ad hoc categories. *Memory & Cognition*, 11(3), 365-379.
- Barsalou, L.W. (1985). Ideals, central tendency and frequency of instantiation as determinants of graded structure in categories. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11(4), 629-654.
- Card, S.K. (1982). User perceptual mechanisms in the search of computer command menus. *Proceedings of the Conference on Human Factors in Computer Systems* (pp. 190-196). Gaithersburg, MD.
- Cohen, J., & Cohen, P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.
- Collins, A.M., & Quillian, M.R. (1969). Retrieval time from semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 8, 241-248.
- Collins, A.M., & Quillian, M.R. (1970). Facilitating retrieval from semantic memory: The effect of repeating part of an inference. *Acta Psychologica*, 33, 304-314.
- Green, D.M., & Swets, J.A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Hauptman, A.G., & Green, B.F. (1983). A comparison of command, menu selection and natural-language computer programs. *Behavior and Information Technology*, 2, 163-178.
- Homa, D., & Konrad, K.E.M. (1987). *Construction of semantic space*. Paper presented at the meeting of the Western Psychological Association, Long Beach, CA.
- Konrad, K.E.M. (1988). *Conjoin: A program to fit two multidimensional spaces together based on a common category*. Unpublished manuscript.
- Kruskal, J.B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1), 1-27.
- Lee, E.S., & MacGregor, J.N. (1985). Minimizing user search time in menu retrieval systems. *Human Factors*, 27, 157-162.
- MacGregor, J., Lee, E., & Lam, N. (1986). Optimizing the Structure of Database Menu Indexes: A Decision Model of Menu Search. *Human Factors*, 28(4), 387-399.

- McDonald, J.E., Stone, J.D., & Liebelt, L.S. (1983). Searching for items in menus: The effects of organization and type of target. *Proceedings of the Human Factors Society 27th Annual Meeting* (pp. 834-837). Norfolk, VA.
- Miller, D.P. (1981). The depth/breadth tradeoff in hierarchical computer menus. *Proceedings of the Human Factors Society 25th Annual Meeting* (pp. 296-300). Rochester, NY.
- Pierce, B.J. (1989). *Optimizing computer command menus: The effects of semantic organization, number of alternatives and omission probability*. Unpublished doctoral dissertation, Arizona State University, Tempe, AZ.
- Polzella, D.J., Hubbard, D.C., Brown, J.E., & McLean, H.C. (1987). *Aircrew training devices: Utility and utilization of advanced instructional features (Phase IV - Summary Report)* (AFHRL-TR-87-21, AD-A188 418). Williams AFB, AZ: Operations Training Division, Air Force Human Resources Laboratory.
- Rips, L.J., Shoben, E.J., & Smith, E.E. (1973). Semantic distance and the verification of semantic relations. *Journal of Verbal Learning and Verbal Behavior*, 12, 1-20.
- Rosch, E.H. (1973). On the internal structure of perceptual and semantic categories. In T. E. More (Ed.), *Cognitive development and the acquisition of language*. New York: Academic Press.
- Rosch, E.H. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104, 192-223.
- Schiffman, S.S., Reynolds, M.L., & Young, F.W. (1981). *Introduction to multidimensional scaling*. Orlando: Academic Press.
- Semple, C.A., Cotton, J.C., & Sullivan, D.J. (1981). *Aircrew training devices: Instructional support features* (AFHRL-TR-80-58, AD-A096 234). Wright-Patterson AFB, OH: Logistics and Technical Training Division, Air Force Human Resources Laboratory.
- Shoben, E.J. (1976). The verification of semantic relations in a same-different paradigm: An asymmetry in semantic memory. *Journal of Verbal Learning and Verbal Behavior*, 15, 365-379.
- Smith, E.E. (1978). Theories of semantic memory. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes* (Vol. 6, pp. 1-56). Hillsdale, NJ: Erlbaum.
- Smith, E.E., Balzano, G.J., & Walker, J. (1978). Nominal, perceptual and semantic codes in picture categorization. In J. W. Cotton & R. L. Klatzky (Eds.), *Semantic factors in cognition* (pp. 137-168). Hillsdale, NJ: Erlbaum.
- Smith, E.E., Shoben, E.J., & Rips, L.J. (1974). Structure and processing in semantic memory: A featural model for semantic decisions. *Psychological Review*, 81, 214-241.
- Snowberry, K., Parkinson, S., & Sisson, N. (1983). Computer display menus. *Ergonomics*, 26, 699-712.
- Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*. New York: Academic Press.
- Tydemann, J., Lipinski, H., Adler, R., Nyhan, M., & Swimpfer, L. (1982). *Teletext and videotex in the United States*. New York: McGraw Hill.

APPENDIX: MODEL PARAMETER, RMSE, AND CORRELATION VALUES
FOR METHOD A AND B BEST-FIT SOLUTIONS

Table A-1. Method A Model Solutions for 7 Correct Alternative Present Equations

Omissn Prob	Hierarchy - Level III			
	8 alts (menu set 1)	4 alts (Nesting close - set 2)	4 alts (Nesting dist - set 3)	2 alts (menu set 4)
0%	M1 = 1.385	M1 = 1.320	M1 = 1.442	M1 = 1.333
	SD1 = .319	SD1 = .337	SD1 = .283	SD1 = .413
	M2 = .845	M2 = .900	M2 = .858	M2 = .839
	SD2 = .280	SD2 = .264	SD2 = .322	SD2 = .288
	C0 = 1.036	C0 = 1.033	C0 = 1.121	C0 = 1.133
	C1 = .658	C1 = .436	C1 = .552	C1 = .516
	RMSE = .080	RMSE = .150	RMSE = .082	RMSE = .107
12.5%	M1 = 1.385	M1 = 1.335	M1 = 1.470	M1 = 1.390
	SD1 = .319	SD1 = .334	SD1 = .277	SD1 = .411
	M2 = .845	M2 = .933	M2 = .894	M2 = .871
	SD2 = .280	SD2 = .256	SD2 = .323	SD2 = .286
	C0 = .935	C0 = 1.010	C0 = 1.104	C0 = 1.098
	C1 = .572	C1 = .418	C1 = .471	C1 = .474
	RMSE = .105	RMSE = .131	RMSE = .070	RMSE = .045
25%	M1 = 1.385	M1 = 1.340	M1 = 1.469	M1 = 1.397
	SD1 = .319	SD1 = .336	SD1 = .283	SD1 = .419
	M2 = .845	M2 = .910	M2 = .910	M2 = .904
	SD2 = .280	SD2 = .242	SD2 = .334	SD2 = .280
	C0 = .968	C0 = .991	C0 = 1.107	C0 = 1.105
	C1 = .695	C1 = .420	C1 = .557	C1 = -.202
	RMSE = .082	RMSE = .086	RMSE = .058	RMSE = .067
37.5%	M1 = 1.385	M1 = 1.326	M1 = 1.464	M1 = 1.413
	SD1 = .319	SD1 = .337	SD1 = .282	SD1 = .413
	M2 = .845	M2 = .905	M2 = .902	M2 = .923
	SD2 = .280	SD2 = .265	SD2 = .280	SD2 = .303
	C0 = .916	C0 = .988	C0 = 1.069	C0 = 1.109
	C1 = .711	C1 = .627	C1 = .729	C1 = -.496
	RMSE = .055	RMSE = .049	RMSE = .012	RMSE = .046

Table A-1. (Concluded)

Hierarchy - Level V			
8 alts (menu set 5)	4 alts (Nesting close - set 6)	4 alts (Nesting dist - set 7)	2 alts (menu set 8)
M1 = 1.295	M1 = .976	M1 = 1.770	M1 = .797
SD1 = .449	SD1 = .351	SD1 = .379	SD1 = .356
M2 = .666	M2 = .611	M2 = .680	M2 = .545
SD2 = .252	SD2 = .171	SD2 = .255	SD2 = .180
C0 = .848	C0 = .748	C0 = 1.137	C0 = .745
C1 = .536	C1 = .485	C1 = .508	C1 = .447
RMSE = .101	RMSE = .109	RMSE = .016	RMSE = .091
M1 = 1.315	M1 = 1.011	M1 = 1.809	M1 = .847
SD1 = .437	SD1 = .348	SD1 = .371	SD1 = .352
M2 = .731	M2 = .613	M2 = .700	M2 = .532
SD2 = .256	SD2 = .184	SD2 = .266	SD2 = .190
C0 = .811	C0 = .719	C0 = 1.030	C0 = .732
C1 = .528	C1 = .442	C1 = .534	C1 = .387
RMSE = .093	RMSE = .113	RMSE = .009	RMSE = .083
M1 = 1.342	M1 = 1.008	M1 = 1.844	M1 = .880
SD1 = .407	SD1 = .360	SD1 = .362	SD1 = .368
M2 = .782	M2 = .629	M2 = .722	M2 = .547
SD2 = .267	SD2 = .192	SD2 = .274	SD2 = .202
C0 = .904	C0 = .735	C0 = 1.095	C0 = .763
C1 = .698	C1 = .460	C1 = .710	C1 = .403
RMSE = .043	RMSE = .096	RMSE = .008	RMSE = .083
M1 = 1.403	M1 = 1.036	M1 = 1.850	M1 = .935
SD1 = .424	SD1 = .371	SD1 = .369	SD1 = .378
M2 = .883	M2 = .632	M2 = .754	M2 = .552
SD2 = .286	SD2 = .198	SD2 = .291	SD2 = .223
C0 = .975	C0 = .768	C0 = 1.134	C0 = .811
C1 = .774	C1 = .500	C1 = .736	C1 = .426
RMSE = .034	RMSE = .073	RMSE = .010	RMSE = .058
Legend. M1 = Di distribution M. SD1 = Di distribution SD. C0 = rapid rejection criterion. M2 = Dc distribution M. SD2 = Dc distribution SD. C1 = rapid selection criterion.			

Note. Free parameters: C0 and C1. Fixed parameters: Number of Alternatives, Omission Probability, M1, M2, SD1, and SD2.

Table A-2. Method B Model Solutions for 7 Correct Alternative Present Equations

Omissn Prob	Hierarchy - Level III			
	8 alts (menu set 1)	4 alts (Nesting close - set 2)	4 alts (Nesting dist - set 3)	2 alts (menu set 4)
0%	M1 = 1.385	M1 = 1.320	M1 = 1.442	M1 = 1.333
	SD1 = .130	SD1 = .138	SD1 = .083	SD1 = .087
	M2 = .845	M2 = .900	M2 = .858	M2 = .839
	SD2 = .281	SD2 = .064	SD2 = .256	SD2 = .159
	C0 = 1.211	C0 = 1.127	C0 = 1.319	C0 = 1.205
	C1 = .718	C1 = .837	C1 = .697	C1 = .726
	RMSE = .028	RMSE = .011	RMSE = .032	RMSE = .005
12.5%	M1 = 1.385	M1 = 1.335	M1 = 1.470	M1 = 1.390
	SD1 = .191	SD1 = .120	SD1 = .095	SD1 = .265
	M2 = .845	M2 = .933	M2 = .894	M2 = .871
	SD2 = .180	SD2 = .155	SD2 = .325	SD2 = .186
	C0 = 1.025	C0 = 1.117	C0 = 1.294	C0 = 1.055
	C1 = .728	C1 = .764	C1 = .610	C1 = .667
	RMSE = .010	RMSE = .021	RMSE = .025	RMSE = .015
25%	M1 = 1.385	M1 = 1.340	M1 = 1.469	M1 = 1.397
	SD1 = .131	SD1 = .177	SD1 = .106	SD1 = .158
	M2 = .845	M2 = .910	M2 = .910	M2 = .904
	SD2 = .315	SD2 = .142	SD2 = .342	SD2 = .180
	C0 = 1.159	C0 = 1.041	C0 = 1.275	C0 = 1.085
	C1 = .745	C1 = .737	C1 = .659	C1 = .615
	RMSE = .014	RMSE = .019	RMSE = .026	RMSE = .012
37.5%	M1 = 1.385	M1 = 1.326	M1 = 1.464	M1 = 1.413
	SD1 = .219	SD1 = .164	SD1 = .196	SD1 = .192
	M2 = .645	M2 = .905	M2 = .902	M2 = .923
	SD2 = .302	SD2 = .272	SD2 = .366	SD2 = .203
	C0 = .996	C0 = 1.086	C0 = 1.145	C0 = 1.092
	C1 = .749	C1 = .696	C1 = .700	C1 = .583
	RMSE = .017	RMSE = .012	RMSE = .009	RMSE = .011

Table A-2. (Concluded)

Hierarchy - Level V			
8 alts (menu set 5)	4 alts (Nesting close - set 6)	4 alts (Nesting dist - set 7)	2 alts (menu set 8)
M1 = 1.295	M1 = .976	M1 = 1.770	M1 = .797
SD1 = .307	SD1 = .174	SD1 = .279	SD1 = .177
M2 = .666	M2 = .611	M2 = .680	M2 = .545
SD2 = .124	SD2 = .100	SD2 = .252	SD2 = .047
C0 = .900	C0 = .770	C0 = 1.276	C0 = .634
C1 = .644	C1 = .573	C1 = .525	C1 = .530
RMSE = .022	RMSE = .030	RMSE = .006	RMSE = .026
M1 = 1.315	M1 = 1.011	M1 = 1.809	M1 = .847
SD1 = .276	SD1 = .086	SD1 = .354	SD1 = .040
M2 = .731	M2 = .613	M2 = .700	M2 = .532
SD2 = .090	SD2 = .184	SD2 = .266	SD2 = .152
C0 = .837	C0 = .853	C0 = 1.031	C0 = .789
C1 = .698	C1 = .524	C1 = .537	C1 = .463
RMSE = .018	RMSE = .020	RMSE = .009	RMSE = .029
M1 = 1.342	M1 = 1.008	M1 = 1.844	M1 = .880
SD1 = .288	SD1 = .176	SD1 = .347	SD1 = .113
M2 = .782	M2 = .629	M2 = .722	M2 = .547
SD2 = .202	SD2 = .081	SD2 = .374	SD2 = .064
C0 = .959	C0 = .737	C0 = 1.225	C0 = .680
C1 = .789	C1 = .592	C1 = .694	C1 = .522
RMSE = .010	RMSE = .026	RMSE = .003	RMSE = .011
M1 = 1.403	M1 = 1.036	M1 = 1.850	M1 = .935
SD1 = .255	SD1 = .156	SD1 = .280	SD1 = .062
M2 = .883	M2 = .632	M2 = .754	M2 = .552
SD2 = .110	SD2 = .141	SD2 = .491	SD2 = .136
C0 = 1.011	C0 = .823	C0 = 1.376	C0 = .849
C1 = .906	C1 = .588	C1 = .705	C1 = .511
RMSE = .004	RMSE = .008	RMSE = .006	RMSE = .007
Legend. M1 = Di distribution M. M2 = Dc distribution M. SD1 = Di distribution SD. SD2 = Dc distribution SD. C0 = rapid rejection criterion. C1 = rapid selection criterion.			

Note. Free parameters: C0, C1, SD1, and SD2. Fixed parameters: Number of Alternatives, Omission Probability, M1 and M2.

Table A-3. Method A Summary Statistics for Predicted Versus Observed Proportions
Across 32 Cells--7 Correct Alternative Present Equation Solutions

Equation	Predicted Values		Observed Values		\bar{n}	r	RMSE
	\bar{M}_i	SD	\bar{M}	SD			
1. Proportion of self-terminating searches resulting in hits	.158	.094	.228	.097	32	.913***	.080
2. Proportion of self-terminating searches resulting in commission misses	.059	.065	.017	.022	32	.849***	.063
4. Proportion of exhaustive searches resulting in hits	.282	.136	.403	.160	32	.867***	.144
5. Proportion of exhaustive searches resulting in commission misses	.054	.033	.039	.024	32	.437*	.034
7. Proportion of exhaustive searches resulting in omission misses	.113	.053	.059	.058	32	.709***	.068
14. Proportion of redundant searches resulting in hits	.082	.043	.054	.054	32	.633***	.050
15. Proportion of redundant searches resulting in commission misses	.065	.041	.012	.015	32	.625***	.062
Across all equations combined	.116	.106	.116	.156	224	.887	.079

* $p < .05$.

** $p < .01$.

*** $p < .001$.

Table A-4. Method B Summary Statistics for Predicted Versus Observed Proportions Across 32 Cells--7 Correct Alternative Present Equation Solutions

Equation	Predicted Values		Observed Values		n	r	RMSE
	M	SD	M	SD			
1. Proportion of self-terminating searches resulting in hits	.233	.097	.228	.097	32	.999***	.008
2. Proportion of self-terminating searches resulting in commission misses	.011	.023	.017	.022	32	.943***	.010
4. Proportion of exhaustive searches resulting in hits	.408	.163	.403	.160	32	.999***	.007
5. Proportion of exhaustive searches resulting in commission misses	.013	.013	.039	.024	32	.199	.036
7. Proportion of exhaustive searches resulting in omission misses	.070	.052	.059	.058	32	.979***	.016
14. Proportion of redundant searches resulting in hits	.063	.054	.054	.054	32	.968***	.016
15. Proportion of redundant searches resulting in commission misses	.015	.016	.012	.015	32	.638***	.013
Across all equations combined	.116	.159	.116	.156	224	.994	.018

*p < .05.

**p < .01.

***p < .001.